

Obstacles of nomenclature

Most disciplines know how to handle the naming of newly discovered objects. Not so the molecular biologists, whose profligate and undisciplined labelling is hampering communication.

There is nothing that impedes comprehension as much as unfamiliar words, but when things are discovered they need to be named. Time was when discoverers were rewarded by the incorporation of their own names into the language of science: the ampere, the dalton, Le Chatelier's principle, and so on. Usefully, eponymous labels can carry additional resonance by their association with the meaning of what was discovered, as with bosons, which display statistical characteristics formulated by Bose and Einstein. But such encapsulated history can be oversimplified — the particle physicists' 'Higgs' should be the Schwinger–Anderson–Englert–Brout–Higgs–Guralnik–Hagen–Kibble–Weinberg–Salam.

Perhaps that inability to identify one discoverer characterizes contemporary science. At the same time, the lack of a common classical education explains today's avoidance of the archaic but otherwise constructive habit of giving new things names that have a Latin or Greek etymology. The consequence has been a descent into whimsy. Murray Gell-Mann started it all with his quark, and that label's successors — strangeness, charm, colour, top and bottom — also do nothing for comprehension.

Last week saw progress in nomenclature by two disciplines. Anatomists banned from professional discourse such homely terms as 'Adam's apple' in favour of internationally recognizable names, while the International Union of Pure and Applied Chemistry at last — after years of controversy — agreed the names of elements 101–109 (see page 10). Both sets of decisions were thoroughly traditionalist in their dependence on classical languages (anatomy) and names of key people and laboratories (elements).

Regrettably, molecular biologists have followed the particle physicists' whimsy with obscurantist enthusiasm. For example, even a knowledge of Shakespeare is no help at all in associating the genes *miranda* and *prospero* with asymmetric cell division — nor, to make

matters worse, is the whimsicality consistent: *The Tempest* has not supplied the equally unassociative names of other genes involved: *numb*, *inscrutable*....

But whimsy is not the only way in which molecular biologists hamper comprehension. Protein labels are nothing less than promiscuous in their ability to switch allegiance without notice or conscience. Witness the shift of 'p21' from p21^{ras} to p21^{waf1}. As the former, p21 was a macromolecule associated with a cascade of signals from receptors at cell surfaces to the nucleus, stimulating cell division. Now (whenever "now" began), p21 refers to a different protein that inhibits the cell cycle. What is worse, it is also referred to as WAF1, CIP1, SDI1 and CAP20. Such problems are pervasive — see also the regulatory protein somehow involved in cell death, known variously as FLIP, Casper, FLAME, CASH and I-FLICE.

The chemists and anatomists are like virtually every other discipline in the methodical way in which their committees treat nomenclature. Such bodies earnestly apply strict criteria: according to the International Astronomical Union, one can name valleys on Mercury only after radio observatories, while planetary features cannot be named after politicians, military or religious figures or contemporary philosophers. In the Earth sciences, minerals can be given names only if they have been discovered in the wild.

Molecular biologists have a long way to go. Some progress is evident — for example, with receptors of cell-signalling molecules known as chemokines, three individuals have become accepted as the joint clearinghouse of nomenclature. Less productively, one hears often of biologists at conferences arguing fruitlessly through the night over which of their protein names should be accepted. One committee cannot clean up molecular biological terminology. But access to and communication within that discipline will be greatly hindered unless more systematic and comprehensible systems of nomenclature are developed. □

Escape from UNEP?

Parties to the United Nations biodiversity convention have been given a rare opportunity.

The executive secretary of the United Nations Convention on Biological Diversity is a man not to be envied. He is employed in Montreal, Canada, to carry out the wishes of no fewer than 169 countries who are parties to the convention. But he may not communicate with any directly. That privilege goes to his employer, the United Nations Environment Programme (UNEP) in Nairobi, Kenya.

Ever since the convention entered into force in 1994, this unhappy arrangement has led to administrative chaos, and contributed to the premature departure of one previous executive secretary. Last week, the current executive secretary, too, was on the brink of seeking alternative employment. A last-minute intervention from member countries (see page 5) has staved off almost-certain crisis.

Many countries are appropriately outraged, not least as terms for his new contract had been offered without consulting the convention's nine-country governing body in violation of an earlier arrangement to do so. The process will now be repeated. Some are calling for a comprehensive review, spelling out the respective

roles of Nairobi and Montreal within the convention.

There is a strong case for freeing the convention from UNEP's unnecessary and bureaucratic involvement, particularly since the convention now has its own secretariat in Montreal. The Geneva-based secretariat of the climate convention reports directly to the United Nations secretary general, and does not need hand-holding by UNEP. There is no reason to treat biodiversity differently.

The protracted conflict between UNEP and the biodiversity convention must not be allowed to continue. The convention has a major programme of work ahead, including negotiations leading up to the drafting of an international protocol on the safety of genetically modified organisms. Tense talks will not be helped by internal disarray.

And the United States government has yet to ratify the convention. The Clinton administration needs the support of conservative Republicans not yet convinced of the case for further international environmental legislation. Images of a biodiversity convention racked by infighting will only add to their reasons to stay out. □



Relocation of geological survey 'could free funds for scientists'

[WASHINGTON] Fears that a lack of money to hire young scientists is draining the US Geological Survey (USGS) of its intellectual vitality lie behind an instruction to prepare for the relocation of the agency's western headquarters from Menlo Park, California, to a site outside the San Francisco Bay area, claims a senior survey official.

A directive from the interior secretary, Bruce Babbitt, asked the USGS to vacate two buildings in Menlo Park, Silicon Valley, and to carry out a speedy review of the costs and consequences of moving most or all of its 810 staff from the site, where \$14 million of the \$80 million budget is being paid in rent.

Babbitt's request for the review stemmed from his desire to spend money on "building the intellectual capital of the organization", rather than on ever-increasing rents, says Tom Casadevall, the survey's western regional director.

The issue of high costs at Menlo Park is not new: in the early 1980s, the survey's national mapping and water resources divisions decided to move staff to other locations. As a result of relocations and retirements, these two divisions now have about 250 fewer people at Menlo Park than in 1982.

Apart from the high rent bill, \$50 million of the total Menlo Park budget of \$80 million is spent on salaries. "This doesn't leave a lot for equipment, fieldwork and so on," says George Hargrove, acting chief of programme support for the western region. A recent rent increase was "an indication of what direction we're headed. When you have overhead eating into the operating budget, alarm bells go off".

Scientists at Menlo Park were stunned by



Seismic shift? The USGS has been asked to plan for the relocation of most activities from Menlo Park.

the suddenness of the request, however. It was conveyed to Casadevall in a memorandum from Gordon Eaton, director of the USGS, released to staff on 25 August and then leaked to the press. As well as demanding a review and "preliminary action plan" by 25 September, Eaton asks for the immediate cancellation of all lease renewals at Menlo Park, and for two buildings (7 and 8) to be vacated within a year. The memorandum sets a five-year timescale for relocating all or most of the other Menlo Park activities.

Buildings 7 and 8 are occupied by the survey's 260-strong earthquake hazards team, together with some other staff. The buildings are leased from a private company, which tried to increase the annual rent from \$2.5 million to more than \$4 million when the lease expired in June.

Casadevall is trying to reassure his colleagues that the dates mentioned are for guidance only, and are not inviolable. They are "intended to impart a sense of urgency", he says. Casadevall adds that some people will not work at Menlo Park because of local house prices. The interior department's

main desire, he says, is to attract more young scientists. "The mean age of [scientists in] the geologic division is 54," he says.

But, whereas Casadevall says that a move to lower-cost premises would release money for building the survey's "intellectual capital", others see a move out of the Bay area as a hazard to its intellectual health. Some worry about the practicality of scientists moving their families, while others cite the usefulness of close collaboration with university scientists at Stanford and Berkeley, and with state and federal agencies in the Bay area. Also, activities such as the earthquake hazards reduction programme and the San Francisco Bay and Delta ecosystem programme use the area itself as a natural laboratory.

Babbitt is said to favour a move to Sacramento, where other parts of the interior department have offices, or to Davis, its neighbouring university town. Having removed scientists from the other branches of the interior department to form the National Biological Service — now the Biological Resources Division of the USGS — Babbitt believes that the USGS should serve as a "science service agency" for its sister bureaus, providing scientific input for decisions about the management of federal land.

But others point out that the survey does congressionally mandated work — including the earthquake hazards programme and water quality monitoring — that is of national importance well beyond the interior department. Local congresswoman Anna Eshoo (Democrat, California) reminded Babbitt of this wide constituency in a sharply worded letter expressing her concern.

Casadevall says that, "whatever's going to happen, it isn't going to happen quickly", and other officials promise that the costs of a move will be considered alongside its benefits. But many at Menlo Park remain anxious. One geophysicist says: "Even if this policy is abandoned next week, it would still be of concern to me that our leaders don't understand us."

Laura Garwin

Exonerated researcher settles for \$3m

[WASHINGTON] Bernard Fisher, a cancer researcher at the University of Pittsburgh who sued after he was dismissed and falsely accused of scientific misconduct, has won \$2.75 million in damages and an apology from the university. The National Cancer Institute (NCI) will pay \$300,000 towards his legal fees as part of the settlement.

The agreement ends a battle that began in 1994, when the government charged Fisher with scientific misconduct and the NCI removed him from his job as head of a major breast cancer study. The case was highlighted by hearings held in the US Congress at the time by Representative John Dingell (Democrat, Michigan) (see *Nature* 368, 679; 1994).

Fisher sued the government and the university, arguing that his constitutional rights to due process and scientific freedom had been violated. In February he was exonerated by government investigators of charges that he knowingly published research containing falsified data. The settlement of Fisher's suit, which was to have gone to trial on Tuesday (2 September), was announced by Fisher and the University of Pittsburgh last week.

Fisher, who is 78, said he was "pleased and satisfied" by the settlement. "It was really done for all scientists and not just for me, because the kinds of things that happened to me, all other scientists are vulnerable to," he says. **Meredith Wadman**

Modest increases prevail in Japanese budget bids

[TOKYO] Japanese science agencies are asking for modest budget increases next year, in the knowledge that the finance ministry will no longer support the sharp rises in funds for research they have enjoyed in recent years.

This marks a clear departure from a government plan set last year, which called for an increase in budget for science and technology of more than 50 per cent over the forthcoming five years (see table below). It also shows that the government's recent moves to restructure the nation's ailing finances are taking precedence over the plan.

The Ministry for International Trade and Industry (MITI) has requested an increase of about 5 per cent, while the Science and Technology Agency (STA) asked for only a 1.4 per cent increment. Overall, the increases for the various science-related ministries and agencies are less than half the percentage rises obtained this fiscal year, and, together with the 'scrap and build' philosophy advocated by the Ministry of Finance to encourage greater flexibility in spending, they may force ministries and funding agencies to look hard at the quality and efficiency of their research funding programmes.

Among the more visible items in this year's budget request is a proposal by the STA to set up a genome research centre inside the Institute for Physical and Chemical Research (Riken), where the STA has been supporting efforts in automated genome sequencing and genome research since the late 1980s.

The STA expects the centre to fulfil a co-ordinating role in genetic research in Japan. But some researchers are not impressed, pointing out that earlier efforts in automated genome sequencing at Riken have failed and that most of Japan's genome research is

concentrated in the universities.

The proposed research centre is part of a coordinated effort to bring together research support schemes in information technology, environmental science, and the life sciences at various ministries. MITI is expected to support innovative software products for protein modelling, while the Ministry for Agriculture, Forestry, and Fishery plans to pursue a second, application-oriented, phase of its rice genome project. The Ministry of Health and Welfare has requested increased funding for research on genetic diseases.

Following the same pattern of joint ministry projects, the Ministry of Health and Welfare and the Environmental Agency will reinforce basic and applied research on the public health impact of dioxins. The government revised maximum exhaust standards for dioxins earlier this year, after reports of high dioxin values around Japan's numerous waste-incineration facilities.

Japanese ministries have used the occasion of the International Climate Conference, to be held in Kyoto in December, to request increases for research on global environmental issues. MITI also wants to reinforce government support for renewable energy, efficient furnaces and clean cars.

Earlier this month, after a proposal by the independent Science Council of Japan, the education ministry announced plans to set up centralized research facilities for environmental research and computer science. But the ministry has kept funding requests for both institutions low-key.

The great losers in this year's budget request are large-scale technology development efforts in nuclear energy and space. Nuclear energy research at both MITI and STA has been cut sharply. Earlier this year, Japan's National Space Development Agency faced the prospect of the smallest budget increase for years, and this has already resulted in the redefinition or demise of several projects (see *Nature* 388, 615: 1997).

The STA is also seeking approval to create new administrative positions. These include five new staff each to manage the transition of the troubled Power Reactor and Nuclear Fuel Corporation (PNC) into a new entity and to upgrade the agency's research evaluation scheme. The STA also wants to create a bioethics-related position inside its life science department.

Overall, the budget requests appear to recognize what critics of the country's public research infrastructure have asserted for some time: that Japan should improve the quality and strategic orientation its research, rather than simply expanding funding. **Robert Triandl**

Scripps Institute plan would boost research space

[SAN DIEGO] The Scripps Research Institute is planning a major expansion at La Jolla, California, aiming to consolidate research facilities on one site and to take on more staff.

The proposal has provoked opposition from doctors working at the hospital that would be closed to clear the way for more laboratories. Research space at the site would increase by about 50 per cent under the plan being considered by the its parent board, the Scripps Institutions of Medicine and Science, which also operate six hospitals in the region.

If approved, the institute's expansion could enhance a burgeoning research community in the area — nearby are the Salk Institute; the University of California at San Diego (UCSD); the Burnham Institute (formerly La Jolla Cancer Research Foundation); the Sidney Kimmel Cancer Center; and numerous biotechnology companies.

Richard Lerner, president of the institute, has said the plan will further his goal of making Scripps "the finest research institute in the world".

The institute, which has an annual operating budget of \$160 million, occupies about 500,000 square feet of laboratory space at the La Jolla site and leases about 250,000 square feet elsewhere. Scientists working at these off-site laboratories would be able to move to La Jolla if the research complex were expanded — a process that would take up to five years.

The plan is to close the 170-bed Green Hospital, adjacent to the institute, and combine it with the nearby, 321-bed Scripps Memorial Hospital, across the street from UCSD's new Thornton Hospital. The San Diego region has too many hospital beds and has experienced major cuts in health insurance companies' payments for care.

The proposal has prompted deep concern among the physicians at the renowned Scripps Clinic Medical Group, the 290-doctor team that provides care at the Green Hospital. The close-knit group fears that the new arrangement would detract from its method of treating patients. Dr. Thomas A. Waltz, a neurosurgeon who heads the group, says: "We do not have a knee-jerk reaction, saying: 'No, no, never'. But there are a large number of issues that no one has begun to consider."

But financial pressure is likely to force the move. It would allow the institute to add further to recent recruits, who include Paul Schimmel, a biochemist from Massachusetts Institute of Technology (MIT); Jeff Kelly, a chemist from Texas A&M; Jamie Williamson, a nuclear magnetic resonance specialist, also from MIT; and Williamson's wife Martha Fedor, a biologist from the University of Massachusetts. **Rex Dalton**

Highlights of 1998 budget requests

	1998 request (¥billion)	Per cent change
Science and Technology Agency (STA)		
Total R&D budget requested	745.0	+ 1.4%
Genome research	6.9	+ 171%
Brain research	13.9	+ 40%
Global environmental research	63.7	+ 14%
Space research	183.2	+ 1%
Nuclear fuel cycle R&D	205.4	- 10%
Ministry for Health and Welfare (MHW)		
Total research funds	92.3	+ 1%
Medical genetics	2.5	-
Research on dioxins	1.5	-
Ministry of International Trade and Industry (MITI)		
Total R&D budget request	499.2	+ 5.7%
Genome research	2.1	+ 350%
Public sector/industry cooperation	6.3	+ 262%
Renewable energy	16	+ 36%
Innovative industrial technology	30.5	+ 9%
Patent system upgrade	60	+ 20%
Nuclear power	23	- 8%

Biodiversity boss to stay, averting crisis

[LONDON] The executive secretary of the United Nations (UN) Convention on Biological Diversity was on the verge of leaving his post this week as simmering tensions between the convention secretariat in Montreal and its parent organization, the UN Environment Programme (UNEP) in Nairobi, came to the boil.

Calestous Juma has been persuaded to stay on temporarily, but only after a last-minute intervention by members of the convention's nine-country governing Bureau who met in emergency session on Monday (1 September) with Juma and a senior official from UNEP.

The Bureau's move has saved the convention from crisis — at least for the time being. Juma's departure would have put back the convention's progress on important issues, including negotiations towards a legally binding protocol regulating the safety of genetically modified organisms. His departure might also have further delayed ratification of the convention by the United States. The Bureau's move will now have strengthened his hand in his protracted tussle with UNEP over how the convention should be run.

UN insiders say Juma had decided to leave after a confidential appraisal from a UNEP ombudsman towards the end of his two-year contract criticized his performance. He was offered a one-year extension. He turned down the offer, and prepared to leave at the end of September.

But at Monday's meeting, Bureau members voiced their anger at UNEP's plans to let Juma leave, which they refused to accept on



Juma: staying in place, for the time being.

an executive secretary. "Basically, UNEP will have to go through the whole evaluation process again in full consultation with the Bureau, and extend the executive secretary's contract until the process is finished," says the representative, who declined to be identified.

With the immediate crisis averted, attention will now turn to the reasons why the biodiversity convention faced the loss of its second executive secretary in two years. Juma's threatened departure mirrors that of his predecessor, Angela Cropper, and centres on a dispute between Montreal and Nairobi over Montreal's attempts to win greater autonomy in implementing the decisions of the convention's 169 member countries — a responsibility that currently rests with UNEP.

"Juma is not allowed to communicate directly with countries without the authority of UNEP, yet has to carry out their directives. He also has little control over recruiting staff, or running conferences, which is also done by UNEP. Yet, when things go wrong, he has to carry the can," says one official.

Many countries are keen to be able to con-

duct business with the executive secretary without UNEP looking over their shoulder. Some feel that their success in persuading UNEP to reconsider its decision to let Juma depart represents an important first step in that direction. But legal experts say that, under the convention's rules, UNEP's executive director has complete and final authority to hire and fire whom she likes.

One leading environmental lawyer says the best way to resolve this dispute once and for all is for the convention's annual conference to agree, preferably by consensus, to sack UNEP as the convention's host secretariat. The convention's executive secretary would then be appointed directly by the UN secretary general, in the same way that he appoints the head of the UN's climate convention. Climate convention member countries play a far more direct role than those in the biodiversity convention.

But arriving at a consensus on this issue will not be easy, nor will the task of choosing an alternative secretariat, "as people are not exactly queuing up to do the job," says the lawyer.

UNEP declined to comment. However, sources close to the organization acknowledge that the executive secretary's job "is not an easy one". They add that autonomy from UNEP would not necessarily make the biodiversity convention function more effectively, as UNEP would have to be replaced by another UN supervisory body. "People think that the climate convention is independent and controlled by the parties. But it answers to the UN," says one official.

Ehsan Masood

Animal researchers should 'start talking' to anti-vivisectionists

[LONDON] One of the founders of the Boyd Group, a forum representing the centre-ground in the animals-in-research debate in the United Kingdom, has called for the organization to talk to ardent anti-vivisectionists, possibly opening a line of communication to the extremist Animal Liberation Front (ALF). The suggestion will be discussed at a group meeting next week.

The call was made by Colin Blakemore, professor of physiology at the University of Oxford. Blakemore is president-elect of the British Association for the Advancement of Science, whose annual science festival at the University of Leeds will host the Boyd Group's first public meeting on 9 September.

Reaction from other members of the group to Blakemore's idea has been mixed. Les Ward, director of the moderate animal rights organization Advocates for Animals, and a co-founder of the Boyd Group, says he does not oppose talking to the ALF. But

Ward says he doubts whether dialogue will result in progress given that the ALF is unlikely to renounce violence as a means of stopping animal experiments.

But Mark Matfield of the Research Defence Society, which represents organizations involved in medical research, says he is completely opposed to any contact with the ALF. "I don't think you can have dialogue with an organization that wants to blow you up," he says.

Blakemore, who along with his family has been a target of ALF violence for many years, says he is prepared to "talk to almost anyone" to achieve progress in the debate. "I am perfectly prepared to see the ALF at the table if their attitude is constructive and there is a chance of progress."

Blakemore says the Boyd Group needs to "move forward" and tackle "more contentious" issues. The group so far has agreed to reduce the numbers of animals

used in research within the organizations it represents.

The group has in the past sought the support of mainstream anti-vivisectionist organizations, such as Animal Aid, and the British Union for the Abolition of Vivisection. But, despite keen interest during preparatory discussions, both organizations have said they are not prepared to join unless a total ban on the use of animals is on the agenda.

Blakemore says these organizations could use the group as an opportunity to convince their critics of the strength of their case. But Bob Coombe, scientific director of the Fund for Replacement of Animals in Medical Experiments and a member of the Boyd Group, says wider membership, although important, might also inhibit progress. "One might argue that we could become a talking shop if every single organization is represented," he says.

E.M.

Consortium aims to revive sterile-mosquito project

[HYDERABAD, INDIA] A controversial US-funded mosquito-control project in Delhi which was closed down by the Indian government 20 years ago (*Nature* 256, 355–357; 1975) is now likely to be revived.

International efforts are under way to relaunch a sterile-insect technology (SIT) project, despite scepticism in parts of India's scientific community, with the aim of eradicating *Anopheles stephensi* — the principal vector for urban malaria — from towns in southern India.

SIT relies on the fact that female mosquitoes mate only once. If this single mating occurs with sterilized males, it results in eggs that do not hatch. So releasing sterilized males in numbers far greater than those of the wild male population could eradicate the targeted vector mosquitoes — provided the area is not reinfested by mosquitoes from the surrounding region.

The move to revive the technique is being led by Christopher Curtis of the London School of Hygiene and Tropical Medicine, Alan Robinson, head of the entomology unit of the Vienna-based International Atomic Energy Agency (IAEA), and Vinod Prakash Sharma, who is due to retire shortly as director of India's Malaria Research Centre (MRC) in New Delhi.

The group presented the idea to last month's global meeting on malaria in Hyderabad, sponsored by the Malaria Foundation, which is based in New York. They are also preparing a report on it for India's Department of Biotechnology, which is said to be sympathetic to the project, provided that environmental and biosafety concerns are met.

Both Curtis and Sharma were prominent scientists of the Delhi project that was terminated in 1975 following an uproar in the Indian parliament about reports that the project was part of US efforts to develop yellow fever as a biological weapon.

These allegations were prompted by the fact that project scientists had been genetically manipulating the *Aedes aegypti* mosquitoes which spread yellow fever — a disease that does not exist in India — and were planning a massive release of sterile males of this species.

Curtis denies that anything sinister took place, and says that the project fell victim to an "ill-informed press and political campaign". The new project proposed for southern India will use mosquitoes sterilized by radiation and will not involve genetic manipulation.

Robinson says the IAEA, with backing from the World Health Organization and the



Pointed problem: a male specimen of *Anopheles*.

United States, plans to raise the \$12 million needed to start the project by holding a series of donor meetings, and to start work next year. Meanwhile, the search is on for an Indian scientist to act as local project leader.

Curtis says that towns such as Salem, in the state of Tamil Nadu in southern India, are ideal sites for SIT because *A. stephensi*, the only vector spreading malaria in these towns, is totally absent in surrounding rural areas. "If we get rid of *A. stephensi* in towns using SIT, no replacement is possible and the area can remain free of malaria for ever," he says.

Some scientists are sceptical, however, citing potential problems with cost, efficiency of separation of the sexes, and what they predict would be the temporary nature of any eradication. "This technology is not workable," warns P. K. Rajagopalan, the former director of the Vector Control Research Centre in Pondicherry, who was involved in the Delhi project. "It is a mystery to me why it is coming back after two decades." P. K. Das, current director of the Pondicherry centre, says India does not need high-technology mosquito control. "What we need is political will to implement low-cost techniques that already exist," he says.

It is true that the long-term efficacy of SIT in eradicating mosquitoes has not been demonstrated. But Robinson says that recent success in controlling screw worm, tsetse and Mediterranean fruitflies "gives us hope we can use this technology in human health".

On the question of separation of the sexes — no females can be released because they are the ones that bite — Robinson also argues that techniques for sexing *Anopheles* mosquitoes have been refined. He says he is confident that "the changed political atmosphere in India" will not allow the 1975 episode to be repeated.

And on the third argument, cost, an official from the MRC told the Hyderabad malaria conference: "Delhi spends \$9 million each year on malaria control without any success. Can't we spend this money just once to eliminate the malaria mosquitoes for ever, using SIT?"

K. S. Jayaraman

US patent office withdraws patent on Indian herb

[NEW DELHI] India has forced the US Patent and Trademark Office (PTO) to revoke a contentious patent it granted two years ago to researchers in the United States on the use of powdered turmeric (*Curcuma longa*) for wound healing.

The PTO withdrew the patent on 13 August after a year-long legal battle with India's Council of Scientific and Industrial Research (CSIR), which argued that turmeric, a native Indian plant, had been used for centuries by its people for wound healing, and so lacked the "novelty" criterion required for patenting.

The Indian agency hired a US patent lawyer and spent \$15,000 to fight the case, which it supported with documents ranging from scientific publications to books on home remedies and ancient Ayurvedic texts on Indian systems of medicine.

Indian scientists claim this is the first time that a move in the United States to patent a traditional remedy from the developing world has been successfully overturned. Earlier efforts by an international coalition of environmentalists to get the US patents on products of the neem tree cancelled ended in failure (*Nature* 377, 95; 1995).

The CSIR's director, Ragunath Mashelkar, said the success of the case had far-reaching consequences for the protection of the traditional knowledge base, "not only in India but in other Third World countries". He said the case also highlights the importance of documenting traditional knowledge, to provide evidence of prior knowledge.

The turmeric patent was granted in 1995 to two researchers, Soman K. Das and Harihar Kohli of the University of Mississippi Medical Center. Their six patent claims covered the oral and topical use of turmeric powder to heal surgical wounds and ulcers. Das and Kohli contested CSIR's objections, but the patent office rejected all their claims.

The patenting of traditional remedies from developing countries became a global issue after patents were granted for neem. These patents were upheld because they covered novel processes for increasing the useful life of azadirachtin, neem's active ingredient.

Mashelkar says that India fought the turmeric patent not for financial reasons, but to uphold "national pride" and to dispel unfounded fears that India was incapable of protecting its traditional knowledge base.

In cooperation with other agencies, the CSIR has already launched a programme to analyse 490 medicinal plants and place the information on CD-ROM. This will be made available to European and US patent offices as a reference guide.

K. S. J.

Curbs on B6 highlight dietary dilemmas

[LONDON] Protests from the health-food industry and from some scientists have persuaded almost 100 members of the British parliament to resist proposals to limit over-the-counter sales of vitamin B6. The MPs have signed a motion opposing legislation, now being prepared by the government, that would restrict sales of the vitamin on safety grounds.

The row about the proposed regulation highlights the problems that many governments face in controlling the use of vitamins and other dietary supplements, which are increasingly used medicinally but rarely regulated in the same way as medicines.

The Ministry of Agriculture, Fisheries and Food (MAFF) wrote to MPs last week with information about the vitamin in an attempt to influence their opinion before the vote on the proposed legislation comes up — perhaps next month. The legislation would limit the amount of vitamin B6 on open sale to 10 mg per day. Doses of between 10 and 50 mg per day would be available for sale only under the supervision of a pharmacist, while doses between 50 and 200 mg would be prescription-only.

Vitamin B6 is registered as a drug in the *British National Formulary*, the official drug register, but because of controversy about toxicity the Medicines Control Agency, which is part of the Department of Health (DoH), is reviewing the conditions of the licence it issued for B6.

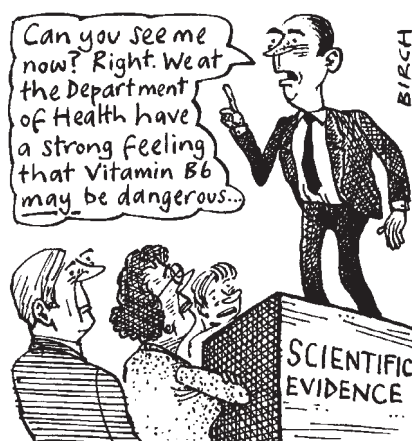
The daily recommended intake for B6 as a vitamin is less than 1.5 mg per day, but higher doses — typically between 50 and 200 mg per day — are widely used to control symptoms of pre-menstrual tension and the side-effects of the contraceptive pill.

No-one disputes that very high doses of vitamin B6 taken over long periods can cause symptoms of peripheral neurotoxicity (nerve damage). But controversy exists about the exact dose that can be deemed safe.

Opponents of the proposed legislation dispute its scientific basis and argue that the public should be free to buy B6 in doses up to 200 mg unless a health risk can be proved.

But the Consumers' Association argues that the scientific basis for concern about the safety of even a 50-mg dose is strong. The association claims the government has mishandled the affair and that this has fed the power of the opposition lobby. Last year, the association asked MAFF to consider the evidence for vitamin-B6 toxicity as part of its general concern that regulation of dietary supplements languishes in a 'no-man's land', falling between the two sets of rules controlling food and medicine.

MAFF referred the matter to the DoH, whose independent Committee on Toxicity in Food, Consumer Products and the



Environment (COT) in July recommended restricting sales. The Royal Pharmaceutical Society of Great Britain endorsed the decision and major supermarkets cleared their shelves of all dietary supplements that included high-dose B6. Meanwhile, the government prepared to draft legislation.

The DoH has not published the full details of COT's scientific deliberations, but a short statement from COT indicates that the committee focused on a 10-year-old study which relies on subjective measures of symptoms. COT acknowledges what it calls the study's "methodological difficulties", but considers it particularly relevant because it concerns patients taking 50-mg doses. COT's statement also refers to its consideration of other studies involving patients taking very high doses of up to 500 mg.

Liz Sheppard, public affairs officer of the Consumers' Association, believes the statement gives a wrong impression of the committee's deliberations. It does not mention, for example, that the committee also referred to well over 600 adverse drug reactions that have been reported by doctors prescribing B6 in the 20 years since it was licensed as a medicine. "This evidence has obviously been

critical in the government's decision," Sheppard says.

The health-food manufacturers' lobby has found it easy to pick holes in the scientific arguments in COT's statement. As well as criticizing the quality of studies quoted by COT, the lobby argues that the safety factor applied by COT, to arrive at its maximum safe dose of 10 mg, is arbitrary. According to COT's rules, safe doses are calculated by determining — usually in animal studies — at what level no adverse symptoms are observed; applying a factor of 10 to compensate for differences between animals and humans; and applying an additional factor of 10 to compensate for variability among individuals.

The manufacturers argue, with the support of some scientists, that applying these factors to essential vitamins is not appropriate. Arnold Beckett, a professor of pharmaceuticals at the University of London and a former president of the Pharmaceutical Society, points out that if the same safety factors were applied to caffeine, supermarkets would have to clear their shelves of coffee.

But Sheppard argues that the whole debate could have been avoided if the DoH had properly explained the scientific basis of its recommendation.

The United Kingdom is not alone in worrying about the possible dangers of B6. The European Union is considering a move to harmonize member countries' systems of dietary supplement control, which vary considerably. In a report on vitamin B6, the union's Scientific Committee for Food has said that "intakes of more than 50 mg per day must... be regarded as potentially harmful".

In the United States, vitamin B6 is freely available for sale, but manufacturers have set a voluntary limit on doses of 200 mg. This may change when a Food and Drug Administration report on dietary supplements is completed in two years' time. **Alison Abbott**

Flood control strategy agreed for Oder

[MUNICH] Germany, Poland and the Czech Republic have agreed on closer collaboration in flood management following the recent catastrophic flooding of the River Oder, which runs through all three countries. Lack of coordination in managing the river has been widely criticized as contributing to the severity of flood damage.

The environment ministers of the three countries and the German state of Brandenburg met last month to determine how best to work together to reduce the consequences of any future floods.

A joint flood management strategy will now be established between the three countries. It will include ecological

measures, such as the protection and creation of flood plains to buffer peak flows, and the development of simulation models for forecasting floods.

Some German university institutes have offered research proposals for the development of flood models to the Brandenburg environment ministry.

The environment ministers called on the help of the International Commission Against Oder Pollution (IKSO), representing Germany, Poland, the Czech Republic and the European Union, which was set up in 1996 with the aim of developing a collaborative strategy for water protection in the region. **Matthias Strobl**

Library tries to rebuild flooded collection

[WASHINGTON] The library at Colorado State University (CSU) at Fort Collins, Colorado, is seeking help from the academic community as it tries to recover from a flash flood that devastated the library's collection.

The flood, which killed five people in Fort Collins, broke through a basement wall of the Morgan Library on the night of 28 July, and rapidly rose to ceiling height. More than 425,000 volumes were being stored in the basement — more than usual, because of renovations taking place elsewhere in the library.

Among the damaged materials were bound volumes of more than 18,000 journals, supporting CSU's strong academic programmes in the sciences, agriculture and forestry.

At least 10 per cent of the damaged volumes were declared unsalvageable in the days following the flood. The rest — some 380,000 volumes — have been sent to the disaster recovery company DRS, Inc. in Fort Worth, Texas, for cleaning, reshaping and freeze-drying.

Although the company estimates that 80 per cent or more of this material can be saved, it will be two years before the last item is returned to CSU — which is too long



After the flood: around 380,000 volumes from the library at Colorado State University have been sent for recovery following the flash flood that breached a basement wall of the Morgan Library.

to wait, according to the librarian's collection development officer, Joel Rutstein. "Twenty per cent is not salvageable, but we don't know which they are. And we can't wait to find out."

"There's no precedent for this," Rutstein says of the loss. No academic library in the United States has lost "the meat of its collection" in this way, he says.

CSU's collection serves an important role for the state at Colorado. "Eleven of our 13

programmes of excellence were affected — we're the only institution in Colorado collecting in those areas," says the dean of libraries, Camila Alire.

CSU estimates the total damage to the campus at \$100–\$125 million, up to half of which is accounted for by damage to the library and its contents. Although the university will recover most of this money from insurance, or from the state and federal governments, much of the library's collection — including back-runs of most journals — cannot easily be re-purchased. The library is accordingly asking for donations of journals and monographs, and has set up a Web site — <http://www.coalliance.org/csuflood/csuflood.shtml> — to coordinate offers.

More immediate help has been forthcoming from other libraries. Many of them have undertaken to give priority to requests from CSU for inter-library loans, and several database services have donated subscriptions.

As for the longer-term prospects, "since we set up the Web site, we've been overwhelmed with responses", Rutstein says. "Although I hate to use the metaphor, offers have gone from a trickle to a tidal wave."

Laura Garwin

Eugenics scandal reveals silence of Swedish scientists

[PARIS] Revelations that successive Swedish governments sterilized more than 60,000 people on eugenic, social and economic grounds over 40 years, and until as recently as 1976, have exposed the failure of the Swedish scientific and medical communities to speak out against the practice.

The scandal broke last week following the publication of a series of articles in the Swedish newspaper *Dagens Nyheter*. Until then, few Swedes knew anything about the sterilization programme, even though its existence had hardly been a state secret. During the 1970s a parliamentary commission debated the 1934 law authorizing the sterilizations. The impact of the law, which was abolished in 1976, has also been widely debated in academic circles recently.

Mattias Tyden, a researcher in the history department at the University of Stockholm and an expert on eugenics, said a broad consensus existed in the scientific and medical communities and across political parties that the introduction of the sterilization programme was a "scientific and modern way of changing society for the better".

Although some individuals questioned the justification for the programme from the outset, their view remained a minority one, said Tyden, who points out that a

parliamentary commission set up during the 1930s concluded that the programme raised no particular ethical issues. Even after the Holocaust, Swedish society "did not see the parallels", he added.

Although the programme had its origins in the eugenics movement — in 1922, Sweden became the first country to create an institute for racial biology — its emphasis quickly shifted away from racial eugenics and towards social and economic goals. Whereas most sterilizations in the 1940s involved mentally handicapped individuals, the majority in the following decades were on women in difficult social circumstances. Some experts say that these sterilizations were carried out to meet these specific circumstances, without any intention of altering the overall composition of the population.

Sterilizations carried out on the basis of eugenic goals peaked in 1944 with 1,437 operations, compared with just 233 on social and economic grounds. But by 1954, sterilizations performed on eugenic grounds had fallen to 204, while those for social and economic reasons had soared to 1,571.

The widely held view was that sterilization was a "human solution" to improving the circumstances of misfits while reducing the social and economic burden they placed on



Swedish legacy of eugenics: Maria Nordin was forced to undergo sterilization in 1938.

society, said Tyden. The law was abolished only after growing opposition from victims and women's movements during the 1960s, and in particular as a result of campaigning in the early 1970s by Karl Grunewald, an inspector at the National Board of Health.

The Swedish scientific and medical community has so far been relatively silent on the subject. The Swedish national medical association declined to comment, claiming that such policy issues fall outside its remit. The Swedish Union of Physicians has not yet taken any position on the issue, according to Elisabeth Frustel, a spokeswoman for the organization, who said that physicians were "working within the law".

Declan Butler

France gives go-ahead for genomics centre on complex diseases

[PARIS] The French government is believed to have approved plans to create a national centre for the genotyping of complex diseases. The centre, which will have a budget of FF50–70 million (US\$8–12 million) annually, will be similar to the UK Wellcome Trust Centre for Human Genetics, which was created in 1994. Mark Lathrop, a leading genome researcher currently at the Wellcome centre, is being tipped to become director of the French centre.

The ministry for national education, research and technology declined to confirm that any decision has been taken, saying that it was “too soon” to discuss details. But, according to reliable sources, it decided last Friday (29 August) to finance the creation of a centre to study the genetics of diseases, such as diabetes, obesity, hypertension and heart disease, that are caused by the interaction of multiple genes with each other and with environmental factors.

The proposal to build the centre, combined with the recent decision to build a French national sequencing centre (see *Nature* 387, 542; 1996), represents an effort by France to regain ground in genomics, having lost its early lead in the mapping of the human genome (see *Nature* 361, 671; 1993). The centre would “create a considerable dynamic” in the study of complex diseases in France, says Florent Soubrier, a researcher at the national biomedical research agency, INSERM, who prepared a report for the government on its proposal.

Japan ends funding for ‘cold’ fusion project

[TOKYO] Japan’s Ministry of International Trade and Industry has decided not to renew a contract for a five-year project on cold fusion when it ends next March.

The ministry has spent around ¥30 billion (US\$25 million) on the cold fusion research at its New Hydrogen Energy laboratory, near Sapporo, since Stanley Pons of the University of Utah and Martin Fleischmann of Southampton University first claimed to have achieved fusion in hydrogen at room temperature in 1989. Along with IMRA, a foundation sponsored by Toyota, the ministry has been the main source of support for cold fusion research internationally.

A ministry official said that the project had succeeded in developing “superior calorimeter technologies for heat excess measurements”. But, even with the assistance of these calorimeters, Japanese researchers — like their colleagues overseas — failed to find heat generated by cold fusion. The

decision to end the work emerged from the ministry’s budget request for 1998, which was released last week (see page 4).

Chemists agree official element names

[LONDON] The International Union of Pure and Applied Chemistry has agreed official names and symbols for nine transfermium elements — those whose atomic number is greater than 100 — resolving disputes about who gains eternal credit for their discovery.

The chosen names, which will be published in the December issue of the journal *Pure and Applied Chemistry* (69, 2471–2473; 1997), are: 101 mendelevium (Md); 102 nobelium (No); 103 lawrencium (Lr); 104 rutherfordium (Rf); 105 dubnium (Db); 106 seaborgium (Sg); 107 bohrium (Bh); 108 hassium (Hs) and 109 meitnerium (Mt).

The names, all of which are already in general use, were formally agreed over several years by a Commission on Nomenclature of Inorganic Chemistry, which had to resolve various disagreements.

France drops plans for public health agency

[PARIS] France has abandoned ambitious plans to create a single public health agency, modelled on the US Food and Drug Administration, that would have been given responsibility for the safety of drugs and foodstuffs, and the quality of air and water.

The proposal for such an agency was included in the election manifesto of the Socialist government elected last June, and the idea has been staunchly defended by junior health minister Bernard Kouchner. But the government has now decided to keep the existing medicines agency, which comes under the jurisdiction of the ministry of health, and to create a separate agency for food safety, jointly controlled by the ministries of health, agriculture and finance.

The decision has disappointed some observers, who wished to see food safety removed from the agriculture ministry altogether in the wake of the ‘mad cow’ crisis. But others argue that the decision nonetheless represents a major improvement in the control of food safety, which is currently the responsibility of a multitude of different ministerial committees.

Task force will inspect safety at nuclear plants

[TOKYO] Japan’s Science and Technology Agency (STA) will appoint a task force to inspect all nuclear facilities operated by the Power Reactor and Nuclear Fuel Corporation (PNC), after last week’s revelation of weak safety and management practices at a radioactive waste storage

facility at the PNC’s Tokai Village plant.

The task force will include independent experts, the STA pledged. The agency also announced plans to contract out analysis of soil and water samples from around the facility to an independent expert team.

The results of the inspection will feed into the deliberations of a working group at STA, headed by Atsuyuki Suzuki, a professor of nuclear engineering at Tokyo University, that is considering the details of the transition to a new corporate structure at the PNC, due to take place later this year.

New Zealand farmer spread rabbit virus

[LONDON] A New Zealand farmer last week admitted to spreading rabbit calicivirus disease (RCD) with the help of other farmers in an attempt to control the country’s rabbit population.

The farmer said that the idea of inoculating rabbits with the virus came from news reports of the spread of RCD among Australian rabbits after it had escaped a quarantined trial in October 1995.

At the end of last month, New Zealand police cordoned off parts of the North Island and South Island areas, set up road blocks and established ‘no-fly’ zones, after the agriculture ministry found dead rabbits at several farms. The rabbits had tested positive for the virus.

The New Zealand government had rejected the use of the virus to control rabbits in July. However, the farmers will not be prosecuted as the deliberate release of organisms is not illegal under New Zealand’s Biosecurity Act.

Remote-sensing NASA satellite loses contact

[WASHINGTON] Spacecraft managers at the US space agency NASA and TRW, Inc. were struggling this week to regain contact with a new satellite that has been out of radio contact since 26 August.

The \$71-million Lewis satellite, which is testing several advances in spacecraft design as well as the feasibility of hyperspectral imagery of Earth from space, was launched into a temporary checkout orbit on 22 August, but ground controllers stopped receiving signals from it four days later. Just before the signal was lost, telemetry indicated that the spacecraft had begun a slow spin, perhaps due to a stuck thruster.

Without its solar arrays pointing at the Sun, power onboard the satellite began to drain away. Mission managers hoped power would return, however, as the Sun’s angle improved early this week. Lewis can stay in its current orbit for about three weeks before atmospheric drag brings it down to burn up in the atmosphere.

Gourmet cannibalism in New Guinea tribe

Sir — June Goodfield is partly right in her defence of Shirley (Glasse) Lindenbaum and Robert Glasse's priority in identifying Fore cannibalism as the aetiological agent in the epidemic spread of kuru¹. The facts are, however, more complicated.

Before February 1966, when Gajdusek, Gibbs and Alpers reported² that kuru is transmissible, speculation about an aetiological agent could be no more than that — speculation. An epidemiologist at the US National Institutes of Health, Leonard Kurland, raised the possibility in a letter to Joseph Smadel, then the institutes' associate director, in 1957 (ref. 3). Gajdusek, who was in New Guinea, as it was then called, reviewed the issue, not for the first time or the last, and rejected the possibility, as he wrote to Smadel, because he could "see no sign of infection or post-infectious phenomena"⁴.

Ann and J. L. Fischer, two American anthropologists, proposed in spring 1961 that "the Fore habit of eating corpses suggests a way in which a viral agent might be passed"⁵. Alpers told me he found such speculation commonplace in bars in Goroka following his arrival in New Guinea in October 1961.

Thus the Glasses' priority is justified not by chronology but by their important role, which Alpers emphasized to me, in collecting detailed evidence of cannibal feasts which could be matched with the subsequent appearance of the disease in participants — evidence first

presented in 1963 (ref. 6).

Both Alpers and Lindenbaum told me that Fore women typically ate every part of the bodies they cannibalized, even the faeces and the bones — the practice was, after all, an alternative burial of family members. Lindenbaum does not attach the word "ritual" to the practice; she calls it gourmet cannibalism. Fore women told her they ate the dead because they were "delicious".

Direct inoculation through the mucous membrane was certainly one probable route of infection, as Gajdusek and Gibbs continue to maintain, but with mountains of cannibal cattle dead of bovine spongiform encephalopathy and sheep of scrapie, it is quibbling at this late date to deny oral transmission of spongiform encephalopathy.

Richard Rhodes

609 Summer Hill Road,
Madison, Connecticut 06443, USA
e-mail: rhodesr@mail.yale.edu

1. Goodfield, J. *Nature* **387**, 841 (1997).
2. Gajdusek, D. C., Gibbs C. J. Jr & Alpers, M. *Nature* **209**, 794–796 (1966).
3. Gajdusek, D. C. (ed.) *Correspondence on the Discovery and Original Investigations on Kuru (Smadel–Gajdusek Correspondence, 1955–1958)*. National Institutes of Health DHEW Publ. No. 76–1168, 253 (1976).
4. Gajdusek, D. C. (ed.) *Correspondence on the Discovery and Original Investigations on Kuru (Smadel–Gajdusek Correspondence, 1955–1958)*. National Institutes of Health DHEW Publ. No. 76–1168, 299–300 (1976).
5. Fischer, A. & Fischer, J. L. *J. Health Hum. Behav.* **2**, 24 (1961).
6. Glasse, R. M. *Cannibalism in the Kuru Region*. (Dept of Public Health, Territory of Papua and New Guinea, Mimeographed, 1963).

nonexistent; nevertheless, the appropriate publicity leads to the fraudster's admission to the Institut de France. It is interesting to come across this comedy just now when scientific misconduct is openly discussed in industrial societies but remains concealed in developing countries.

The small scientific output of Mexico and the lack of high-profile fraudsters do not guarantee proper behaviour among Mexican scientists²; in fact, it may well be the other way round. Developing countries also need to set up procedures to deal with scientific misconduct.

Horacio Rivera

División de Genética,
Instituto Mexicano del Seguro Social,
Ap. Postal 1-3838, Guadalajara, Jalisco, Mexico
e-mail: hriviera@udg.serv.cencar.udg.mx

1. Romain, J. *Donogoo* (Spanish version by Massanes, N. & Losada, S. A., Buenos Aires, 1957).
2. Rivera, H. *Arch. Med. Res.* **27**, 587–588 (1996).

Scientific advance thrives on openness

Sir — I greatly enjoyed your recent leading article on "Flawed understanding of the scientific process", with its thoughtful emphasis on science as a continually evolving path to better understanding rather than a set of certainties (*Nature* **388**, 607; 1997).

Your leading article was set against the background of recent controversies in the United States and in France. I think there are, however, interesting resonances between the principles you set out and those that underpin the UK government's document *The Use of Scientific Advice in Policy Making* (DTI/Pub 2808/0.5k/5/97/RP. <http://www.dti.gov.uk>).

This document was prepared by the Office of Science and Technology for the previous government, and strongly endorsed by the present Minister for Science, Energy and Industry, John Battle, in his speech to the Parliamentary and Scientific Committee on 10 July: "Involve at least some experts from other, not necessarily scientific, disciplines, to ensure that the evidence is subjected to a sufficiently questioning review from a wide ranging set of viewpoints; [and] ensure that data relating to the issue are made available as early as possible to the scientific community to enable a wide range of research groups to tackle the issue. Scientific advance thrives on openness and competition of ideas."

Robert M. May

Chief Scientific Adviser to the UK Government,
Office of Science and Technology, Albany House,
94–98 Petty France,
London SW1H 9ST, UK

Pseudo-authorship

Sir — A recent leading article (*Nature* **387**, 831; 1997) raises the question whether authorship should be redefined. I think it should: readers should know who is really responsible for published research results.

As things stand, researchers need to publish as many articles as possible and this leads to multi-authorship, gift authorship and 'salami' tactics. Journals prescribe how to submit a manuscript, so why not prescribe correct authorship?

In more than 25 years working as a scientific editor (in geology, nuclear energy and technology) and in national and international editorial organizations, I have not been aware of any valid argument for more than three authors per paper, although I recognize that this may not be true for every field.

Perhaps scientific journals and

international organizations could take the lead. If journals were to instruct authors that manuscripts with more than three authors would not normally be considered for publication, there would soon be a drop in the number of pseudo-authors.

A. J. van Loon

R&D Text Consulting, PO Box 336,
6860 AH Oosterbeek, The Netherlands

Fraud foreseen

Sir — The widespread occurrence of fraud and misconduct in scientific research prompts me to honour the foresight of Jules Romain (1885–1972) who satirized such improper behaviour many years ago¹. His farcical comedy *Donogoo* (1920) deals with the description by the famous geographer Yves Le Trouhadec of the golden city of Donogoo-Tonka, which is later shown to be

Achieving low-cost emissions targets

There has been much discussion about ways to reduce the rate of atmospheric accumulation of carbon dioxide and other greenhouse gases. What are the key economic and policy issues concerning the timing of such efforts?

**Stephen H. Schneider and
Lawrence H. Goulder**

In thinking about ways to achieve a reduction in the emissions of greenhouse gases, it is crucial to distinguish between the timing of emissions abatement and the timing of policy action. As we discuss below, there are compelling arguments for implementing only relatively small CO₂ mitigation measures in the short term and delaying significant abatement to the more distant future, but this does not justify the absence of policy action now. On the contrary, it is vital to have a short-term abatement policy to bring about low-cost reductions in CO₂ emissions, even when most of those reductions will occur in the distant future.

We believe that a carbon tax is the most economically efficient and administratively flexible instrument for policy action now. Under a range of plausible economic assumptions, we show here that a carbon tax is a better instrument for generating a cost-effective abatement profile than subsidies to research and development for alternative, low-carbon sources of energy.

Postponing abatement

Last year, Wigley, Richels and Edmonds¹ published a widely cited, influential article on the appropriate timing of emissions abatement. They examined several alternative models for mitigation pathways culminating in the same long-term concentration of CO₂ and concluded that, in general, overall abatement costs are kept to a minimum if the bulk of CO₂ abatement takes place in the more distant future rather than soon.

Wigley *et al.* offered four explanations for their conclusion. The first is the positive return on capital: capital growth would mean that fewer resources need be put aside today to fund future abatement. The second is that the capital stock for energy production and use may be long-lived. Delayed abatement would allow expensive production assets to be gradually replaced with fewer carbon- or energy-intensive new technologies after the older stock has reached the end of its economic life. The third is technological progress: new, energy-efficient technologies will be discovered and developed, hence the "availability of low-carbon substitutes will probably improve and their costs reduce over time"¹. And fourth, carbon emitted sooner is exposed to natural removal processes for longer, so delayed abatement permits a larger total of cumulative emissions to produce the same long-term concentration target.

The main idea expressed by Wigley *et al.*, that society can economize on the costs of achieving targets for reduced emissions of greenhouse gases by deferring most abatement to the future, has won many adherents. But others (see, for example, ref. 2) contend that delay could be costly, as a commitment to reduce emissions now would be an important catalyst for technological progress that will ultimately lead to lower costs of abatement. Much of this disagreement can be overcome by distinguishing between emissions abatement and abatement policy. Although there are strong arguments to justify delaying the bulk of emissions abatement, we believe there are compelling arguments for short-term implementation of government policies to set in motion the economic adjustments necessary to bring about low-cost reductions in emissions.

A carbon tax today

Climate-related damage from accumulation of CO₂ is a cost that, without government intervention, is not incorporated in the market price of fuels³. A carbon tax, if set appropriately, would allow market prices to reflect the full social cost of climate damage and give purchasers the incentive to economize on fuel use. In addition, because the tax would cause prices of carbon-intensive consumer goods to rise relative to other goods, it would give individual consumers the incentive to rely more heavily on less-carbon-intensive goods and services, and producers the incentive to develop alternatives.

Of course, other policies could be used to discourage or restrict the use of carbon-intensive fuels, for example, direct limits or caps, fuel-efficiency standards and mandated technologies or equipment. There have been many economic comparisons of these alternatives^{4,5}, but most tend to embrace the carbon tax as the most economically efficient and flexible instrument because it imposes fewer information requirements on policy-makers; it provides dynamic incentives; is relatively inexpensive to administer; and is relatively easy to adjust in response to new information (essential in climate assessment). Moreover, a carbon tax would bring governments revenues that could be used to finance cuts in ordinary income taxes, thereby helping to avoid inefficiencies associated with disincentives to work or to save. This ability to generate revenue would make a carbon tax much more efficient than limiting carbon use⁶.

An appropriately scaled carbon tax would also help to dictate the most efficient

time-profile for abatement. We agree with Wigley *et al.*¹ that it is more cost-effective to defer to the future the bulk of carbon emissions abatement (relative to the path of emissions under an unconstrained 'business-as-usual' model). But this does not justify avoiding carbon taxes in the short term.

Economic analysis indicates that carbon tax rates should be set according to the 'marginal environmental damage' from CO₂ emissions. Most analyses imply a carbon tax rate that rises with time^{3,7,8} to encourage the level of abatement to increase. Even a constant carbon-tax rate, though less economically efficient, is consistent with a rising time-profile for abatement because, as new technologies for emissions (or fuel) abatement are discovered and implemented, the profit-maximizing amount of emissions-abatement expands, even if the tax rate is constant. Thus, there is no contradiction between introducing carbon taxes now and a cost-effective time-profile for abatement.

Introducing the carbon tax now could be a key factor in inducing the technological change that justifies deferring most abatement to the future. Making carbon-based fuels more expensive provides incentives to research alternative energy-supply options. As emphasized in ref. 2, a carbon tax can also stimulate technological change by promoting "learning-by-doing". In the context of climate change, this means that experience with processes that reduce CO₂ emissions may lead to the discovery of cheaper ways to reduce emissions. In our view, introducing the carbon tax now would usefully exploit this phenomenon.

Research subsidies or carbon taxes?

Does the fact that a carbon tax may stimulate research mean there is no justification for introducing a research and development subsidy as part of climate-change policy? A general economic principle is that governments should apply the policy instrument most closely related to a particular 'market failure'. As noted above, the central market failure in this case is the climate damage associated with combustion of carbon-based fuels. A research subsidy does not directly deal with this market failure because, unlike a carbon tax, it does not directly alter the prices of carbon-based fuels. If there were no other market failures to be concerned about, a subsidy would be unnecessary: the carbon tax alone would reduce the accumulation of CO₂ in a cost-effective manner.

But there *is* a second market failure —

failure in the market for research and development. It is well known that private investment in research and development can generate 'spillover benefits' that are enjoyed by parties other than the investor. Not all knowledge can be kept as private knowledge. Under these circumstances, companies tend to underinvest in research and development, mainly because they do not take into account the full social value (including the spillover benefits) when they make the investments. A government subsidy can correct this market failure, ideally by lowering a company's costs enough to allow it to expand its investment in research and development to the socially optimal level.

Does this argument imply that a carbon tax should be accompanied by a subsidy for research into alternative energy supplies? It depends. Economic theory suggests that a subsidy is appropriate wherever there are significant spillovers. If such spillovers arise in connection with virtually all industrial research investment, then the most economically efficient policy response is a broad-based subsidy for a wide spectrum of industries. On the other hand, if such spillovers are specific to investment in alternative energy supplies, then there is a basis for a more targeted subsidy. Although there are strong theoretical and empirical reasons to support a carbon tax, the case for a targeted research and development subsidy is less clear.

We have developed an economic simulation model for the United States⁹ which we believe is the first large-scale, general equilibrium, economic model to take into account incentives to invest in research and development, knowledge spillovers, and the functioning of research and development markets. The estimated costs of reducing cumulative CO₂ emissions by 15 per cent in the 100 years after 1995 is shown in Table 1. A research subsidy alone never offers the cheapest way to meet the target reduction in cumulative emissions and indeed can be many times more costly than the other policies. Results from our model are sensitive to parameters that are highly uncertain; thus, we emphasize the qualitative pattern from Table 1, not specific numbers. These simulations support our view that a carbon tax is essential for cost-effective reductions of CO₂ emissions, and that this tax should be accompanied by a research subsidy only when there are spillover benefits from research and development.

Two qualifications deserve mention. First, if research and development markets are already highly inefficient (for example, distorted by previous subsidies or taxes), then gauging the costs of new subsidies that should be applied to low-carbon energy sources becomes more complicated. Depending on the array of pre-existing subsidies,

Table 1 Costs of 15% reduction in CO₂ emissions 1995–2095

Model	Carbon tax alone	Targeted R&D subsidy alone	Carbon tax plus targeted R&D subsidy of 10%	Carbon tax plus broad R&D subsidy of 10%
1. No spillover from R&D	0.94	8.52	1.02	1.18
2. Spillovers from R&D only by alternative energy industry	0.66	5.98	0.60	0.78
3. Spillovers from R&D investment by all industries	1.03	9.55	1.09	0.81

Figures are percentage reductions to the present value of GDP. All simulations involve carbon tax rates that increase at a rate of 5 per cent annually to the year 2075 and remain constant thereafter. The carbon tax profile is the lowest path of (rising) tax rates that leads to the 15 per cent reduction in cumulative emissions relative to the baseline model. Model 1 evaluates costs when there are no spillovers from research and development (R&D) investments. In this case, the cheapest way to attain the 15 per cent abatement target is through a carbon tax alone. Model 2 assumes that there are significant spillovers from investments in R&D by the 'alternative energy' industry (the non-carbon-based fuel industry). In this case, the combination of carbon tax and R&D subsidy to alternative energy is the more cost-effective way to attain the target. Model 3 assumes all investments in R&D involve significant spillovers. In this case, the least-cost policy involves a combination of carbon tax and broad subsidy to R&D.

the costs of a combination of carbon tax and subsidy might be higher or lower than suggested in Table 1, but this does not affect the general principle of our argument. Second, our model focuses on private-sector research (which may be publicly subsidized). Thus, our simulations do not directly address the effectiveness of research that is directly undertaken by the public sector.

Political and other considerations

A carbon tax policy is likely to be less popular than subsidies for research into energy technologies, even though a policy consisting only of subsidies would not be the cheapest way to reduce emissions. A subsidy is more favourable to producers but a carbon tax is more attractive to general taxpayers (although some low-income groups are concerned about its potential regressivity). A research subsidy needs to be financed through other tax revenues, whereas a government can 'recycle' carbon-tax revenues to the benefit of general taxpayers (including lower-income groups). Discussions of policy options need to take into account these effects on the tax-paying public.

We have concentrated here on the issues of cost-effectiveness and economic efficiency. Of course, in the political arena there are advocates for other policy principles, such as the precautionary principle or the principle of stewardship. Our purpose is not to debate the relative merits of alternative criteria for policy choices, but simply show that a 'do-nothing' policy is difficult to justify, even according to a policy criterion endorsed by most economists.

A further important issue is the existence of uncertainty. One view is that it is premature to initiate policy action now, given the uncertainties about the extent of climate change and associated damage, whereas another is that postponing policy action would be much more costly in future if action were ultimately required. Recent analyses indicate that it is worthwhile starting policy action now, despite the uncertainties, provided that there are not

prohibitively high 'sunk costs' (unrecoverable one-time costs) of introducing particular policies^{10,11}. A carbon tax policy is attractive because it involves minimal sunk costs, suggesting that short-term action is warranted, and because it is flexible enough to allow adjustments as new information about climate change and the effectiveness of policy becomes available¹².

In conclusion, we support the view that it may be cost-effective to defer most CO₂ abatement, but we believe that policy action is needed now for cost-effective future abatement. A carbon tax (or other flexible, direct policy confronting the effects of fossil-fuel combustion) is an essential element of greenhouse policy. A research and development subsidy could be a useful complement to a carbon tax when there are research and development market failures. But the case for a subsidy (in terms of economic efficiency) rests primarily on spillover benefits from general research (or other research and development market distortions) and not on the prospect of environmental damage from atmospheric build-up of CO₂. □

Stephen H. Schneider and Lawrence H. Goulder are respectively in the Departments of Biological Sciences and of Economics, Stanford University, Stanford, California 94305-5020, USA. Both authors are co-appointed in the Institute for International Studies at Stanford.
e-mail: shs@leland.stanford.edu

- Wigley, T. M. L., Richels, R. & Edmonds, J. A. *Nature* **379**, 240–243 (1996).
- Grubb, M. *Energy Policy* (February 1997).
- Nordhaus, W. D. *Am. Econom. Rev.* **72**, 242–246 (1991).
- Hahn, R. & Stavins, R. *Ecol. Law Quart.* **18**, 1–42 (1991).
- Office of Technology Assessment *Environmental Policy Tools: A User's Guide* (US Govt Printing Office, Washington DC, 1995).
- Goulder, L. et al. *Rand J. Econom.* (in the press).
- Peck, S. C. & Teisberg, T. J. *Energy J.* **13**, 71–91 (1992).
- Goulder, L. H. & Mathai, K. Working Paper, Department of Economics (Stanford Univ., 1996).
- Goulder, L. H. & Schneider, S. H. *Res. Energy Econom.* (submitted).
- Manne, A. & Richels, R. *Buying Greenhouse Insurance* (MIT, Cambridge, Massachusetts, 1996).
- Pindyck, R. Working Paper, Department of Economics (MIT, Cambridge, Massachusetts 1996).
- Schneider, S. H. *Envir. Modelling Assessment* (in the press).

Reversing the kinesin ratchet — a diverting tail

R. A. Cross

Kinesins are molecular motors that are mechanically programmed to move along microtubules in one direction only — towards either the rapidly growing plus ends or the more static minus ends. A remarkable new experiment reverses the directionality using protein engineering.

The kinesins are a family of motor proteins that haul cellular cargo along microtubules in a stepwise fashion. Microtubules have an intrinsic polarity, with so-called 'plus' and 'minus' ends, and microtubule-binding motor proteins 'read' this polarity, and then move towards one end or the other. By laying down an organized array of microtubules and then attaching cellular components to an appropriately directed motor, the cell can sort and segregate its components relative to the polarized array of microtubules: this kind of process is responsible for events such as the spatial separation of two sets of chromosomes at cell division.

All evidence to date indicates that the directionality of the kinesin-family proteins is 'hard-wired' — some members move towards the plus ends of microtubules, whereas others move towards the minus ends. All of these motor proteins have some kind of 'heads on a stalk' architecture, and they move along microtubules through a cyclic, forceful interaction of the heads with the microtubule track. Yet the heads of oppositely directed family members seem to be very similar, both structurally¹ and kinetically^{2,3}. So what determines directionality?

In a remarkable experiment reported on page 93 of this issue, Henningsen and Schliwa⁴ describe how they have forced the heads of a minus-end kinesin, non-claret disjunctional (ncd), to reverse their usual direction of motion by splicing them to the tail of a plus-end kinesin, *Neurospora* kinesin (Nkin). This is a very powerful result. Not only does it reveal how to reprogramme the direction of stepping, but it gives an insight into the stepping mechanism — the mysterious, cyclic, mechanical ratcheting action by which the motor makes progress along the microtubule.

Individual heads of kinesin-family motors can generate directional impulses of force. Each head is an ATPase enzyme (it can hydrolyse ATP to ADP), and an array of heads — tethered through short stalks to a glass microscope coverslip and supplied with ATP — will drive microtubules to slide over their surface. Sustained directional sliding is due to

lots of small, directional hops, which are contributed by the many different heads that interact with the microtubule. So how does the new tail-swap experiment reverse the direction of progress? To think about that, we need to consider two possible ways in which the heads might generate directional force.

In one class of models for directional stepwise progress, the heads attach to the microtubule, then undergo a forceful conformational change (a change in shape or tilt). The progress produced by such a directional conformational change will be modest compared with the 8-nm nearest-neighbour distance between binding sites⁵. But the motion can be amplified by attaching a semi-stiff lever arm^{6,7}. There is electron-microscopic evidence for such a conformational change in kinesin — the release of ADP tilts microtubule-attached kinesin heads towards the plus end of the microtubule. Moreover, as some myosins release ADP, they undergo a conformational change which drives the motion of a lever arm⁸. Experiments showing that myosin moves more rapidly along actin when the lever arm is lengthened by protein engineering⁷ have also been interpreted as supporting the model for a directional conformational change.

To explain Henningsen and Schliwa's result⁴ using this kind of model, oppositely directed lever-arm motions would be required for ncd and Nkin. But this is a possibility because, in ncd, the stalk is attached to the amino terminus of the head, whereas in Nkin the stalk is attached to the carboxy terminus of the head. A more serious problem is that any directional conformational change would be expected to reverse before the head detaches, because detachment occurs when the motor switches back to a conformation analogous to the one in which it initially bound (Fig. 1). Myosin is thought to work in a way that circumvents this difficulty, through strain-dependent chemistry, such that following reversal of the power stroke, the heads detach very rapidly⁹.

For kinesins, these and other difficulties disappear if directional progress is due to an

entirely different process — the directionally biased capture of a diffusing¹⁰ head, in a manner related to that proposed by A. F. Huxley in his classic model for myosin¹¹. The key property of this sort of model is that the lever arm is tensioned by thermal rattling before the motor attaches to the track (top line in Fig. 1). A motor that ratchets by this mechanism works as a kind of sliding clamp and, in principle, it does not require any actively driven lever-arm motion at all. The motor acts as a rattle-rectifier, progressing by selectively capturing thermally driven excursions at the extremes of the energy distribution in the productive direction.

By directionally biasing their diffusional approach to the microtubule, functionally identical motor heads can be made to move in opposite directions¹². Herein, I think, lies the most likely explanation of Henningsen and Schliwa's result — that the Nkin tail confers opposite directionality on the ncd heads by reversing the bias of the diffusion-to-capture process. A diffusional mechanism is

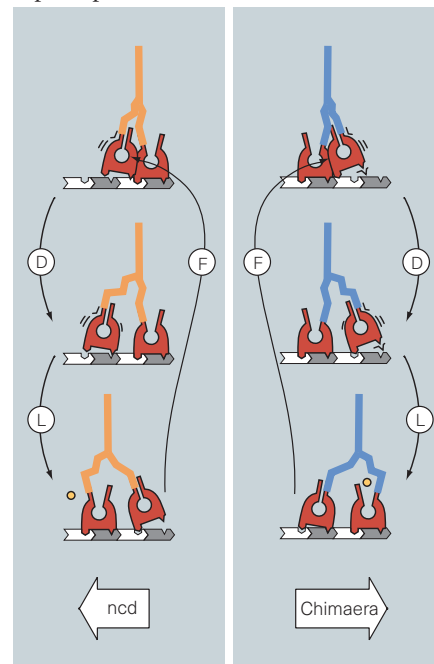


Figure 1 Tail-induced directional bias. By attaching the head of a minus-end kinesin, ncd, to the tail of a plus-end kinesin, Nkin, Henningsen and Schliwa⁴ have created a chimaeric kinesin. Surprisingly, the heads reverse their usual direction of motion, and the chimaera progresses towards the plus end of microtubules. The carboxy terminus of the Nkin tail is attached to the amino terminus of ncd. As discussed here, the tail may act as a spring, recovering detaching heads and causing them to fly forward (process F) to a null point close to the next productive binding site. The Nkin tail (blue) would then display the ncd heads (red) with a plus-end bias. From the null point, the heads diffuse (D) on to the adjacent binding site, and they may or may not make a subsequent lever-arm motion (L) which would then be coupled to microtubule-activated ADP (yellow) release.

plausible¹³ and the amount of bias need not be great, just enough to ensure that binding of the head to counterproductive sites (back-steps) is rare¹⁰.

There is electron microscopic evidence for a directional bias in the head domains of ncd and Nkin. Both motors bind to microtubules through only one of their two heads — the other head projects clear of the microtubule in the direction of its imminent new binding site^{14,15}. The electron microscope images clearly show that the two heads interact only at the point at which they divide, indicating that the directional bias is a function of this bifurcation (head–tail junction) region. This may be why previous attempts to reverse direction using tail-swap chimaeras were unsuccessful¹⁶. The new chimaera makes the join further up into the head, so local structure around the point of bifurcation is probably preserved.

The chimaera does not function perfectly. It runs maximally at only about one-third the speed of the authentic ncd motor. Moreover, microtubules make only short runs when they are sliding on a chimaera-coated surface, tending to detach more easily than from authentic (nonchimaeric) Nkin. So, both the heads and tails could have an intrinsic directionality, in which case the direction of the chimaera would be a result of the ncd heads fighting a losing battle with the Nkin tail. It is

also still a puzzle how single heads are directional¹⁷. Biasing the binding direction by a different mechanism (for example, by having ncd and kinesin/Nkin scan for a binding site using different surface loops³) is one possibility. Like all the best discoveries, the findings of Henningsen and Schliwa beg many more questions than they answer. But they carry the hallmarks of a great experiment — straightforward, definitive, elegant and surprising. □

R. A. Cross is in the Molecular Motors Group, Marie Curie Research Institute, The Chart, Oxted, Surrey RH8 0TE, UK.

1. Kull, F. J., Sablin, E. P., Lau, R., Fletterick, R. J. & Vale, R. D. *Nature* **380**, 550–555 (1996).
2. Sablin, E. P., Kull, F. J., Cooke, R., Vale, R. D. & Fletterick, R. J. *Nature* **380**, 555–559 (1996).
3. Amos, L. A. & Cross, R. A. *Curr. Opin. Struct. Biol.* **7**, 239–246 (1997).
4. Henningsen, U. & Schliwa, M. *Nature* **389**, 93–96 (1997).
5. Block, S. M. *Trends Cell Biol.* **5**, 169–175 (1995).
6. Holmes, K. C. *Curr. Biol.* **7**, 112–118 (1997).
7. Block, S. M. *Cell* **87**, 151–157 (1996).
8. Whittaker, M. et al. *Nature* **378**, 748–751 (1995).
9. Geeves, M. A. *Biochem. J.* **274**, 1–14 (1991).
10. Peskin, C. S. & Oster, G. *Biophys. J.* **68**, (Suppl.) 202–210 (1995).
11. Huxley, A. F. *Prog. Biophys.* **7**, 225–318 (1957).
12. Lockhart, A. & Cross, R. A. *EMBO J.* **13**, 751–757 (1994).
13. Howard, J. *Annu. Rev. Physiol.* **58**, 703–729 (1996).
14. Hirose, K., Lockhart, A., Cross, R. A. & Amos, L. A. *Proc. Natl Acad. Sci. USA* **93**, 9539–9544 (1996).
15. Arnal, I., Metoz, F., DeBonis, S. & Wade, R. H. *Curr. Biol.* **6**, 1265–1270 (1996).
16. Stewart, R. J., Thaler, J. P. & Goldstein, L. S. *Proc. Natl Acad. Sci. USA* **90**, 5209–5213 (1993).
17. Inoue, Y. et al. *Proc. Natl Acad. Sci. USA* **94**, 7275–7280 (1997).

Mathematics

What you eat is how many you are

Ivar Ekeland

How do people make decisions? A standard answer (not the only one) runs as follows: every individual first ranks all alternatives open to them, and then chooses the top of the list. This model had not been validated by econometric studies, but now a simple realization has changed the situation: two people don't act the same way as one.

This picture of decision-making can very easily be cast into a mathematical model. If K is the set of possible decisions, the ranking is defined by the 'utility function', $U(K)$. The individual prefers one decision over another if the value of its utility function is higher. Taking the right decision then amounts to finding in the set K the point x that yields the greatest utility $U(x)$. In this way, every problem in decision-making can be translated into a mathematical problem.

In the case of buying goods, the decisions to be made are of a very restricted scope. The individual has an amount of money w , and we watch them go shopping. If there are N products available in the market, the individual might decide to consume quantity x^1 of product 1, quantity x^2 of product 2, and quantity x^N of product N . If the price of prod-

uct i is p_i , this decision would cost $p_1x^1 + \dots + p_Nx^N$. The consumption profile is affordable only if this sum is not more than w . That constraint defines the set K of feasible decisions, and the individual seeks to maximize his utility $U(x^1, \dots, x^N)$ over that set.

This is called the *Homo oeconomicus* model. Much has been built on these theoretical foundations — in fact, almost everything we know about the way markets operate and how people behave under competition. But do people really behave that way? Is *Homo oeconomicus* anything but a very distant relative of *Homo sapiens*? What is needed, as with any disputed theory, is an experimental test.

Devising such an experiment is not easy. Utility functions are like unicorns: even if they exist, no one has ever seen one. I do not know what my utility function is, and I would be hard put to answer any questions about it. On the other hand, even if I do not know how I make decisions, I certainly do make them, and the result is observable: what I buy is there for all to see. Hence a first question: is there anything to be said about my consuming behaviour that would be true if the *Homo oeconomicus* model is correct, and false otherwise?

Of course, there is little to be gained by observing just one decision: the trick is to vary the price system and to follow the response. In other words, observe the changes in my consuming pattern as prices vary: this is called the individual demand function. If the *Homo oeconomicus* model is correct, this amounts to following the solution $(x^1, \dots, x^N)$ of the optimization problem as a function of the parameters $(p_1, \dots, p_N)$.

We then run into two interesting mathematical problems. First, given an observed demand function $x(p)$, how can we tell if it is compatible with *Homo oeconomicus*? That is, are there special properties that the demand functions should satisfy if they arise as solutions of a minimization problem? And if so, can one recover the utility function of the individual from his demand function?

Solutions to the minimization problem have to satisfy a system of nonlinear partial differential equations called the Slutsky relations^{1,2} (see box). This provides us with the test we were looking for: observe the demand function of an individual, and check whether it satisfies the Slutsky relations. If not, then the *Homo oeconomicus* model should be discarded.

Unfortunately, every econometric study until 1995 has found the Slutsky relations to fail. Of course, there is much at stake, and a lot of effort has gone into explaining why one should not discard the *Homo oeconomicus* model, in spite of the fact that it is appar-

Shopping for pairs

According to the *Homo oeconomicus* model, people make purchasing decisions by trying to maximize a utility function $U(x^1, x^N)$, where x^i is the amount of each commodity. With a fixed total amount of money w , and prices p_i , then $p_1x^1 + \dots + p_Nx^N \leq w$.

●The Slutsky relations

Consider the matrix $S \equiv s^{ij}$ defined by:

$$s^{ij} \equiv \partial x^i / \partial p^j + (\partial x^i / \partial w) x^j$$

If the choice $x(p, w)$ maximizes U , then S must be symmetric:

$$s^{ij} - s^{ji} \equiv 0$$

●The Browning–Chiappori relations

For a two-person household, the matrix S is instead constrained solely by efficiency, and is the sum of two matrices, one symmetric and the other of rank one:

$$s^{ij} - s^{ji} \equiv a^i b^j - a^j b^i$$



Figure 1 Tyrone Power tries to work out what has happened to his demand function. (From *That Wonderful Urge*, 20th Century Fox.)

ently not supported by the data. The latest explanation is the most startling: M. Browning and P. A. Chiappori claim³ that the *Homo oeconomicus* model is correct: it is just not being taken seriously enough.

They rightly point out that none of the econometric studies conducted to check the Slutsky relations discriminate between individuals and households. However, *Homo oeconomicus* most definitely is single. As soon as two people pool their resources, they will have different utilities. The decision-making process then becomes a bargaining process, and there is no reason to expect that the outcome could be described by a single utility function. In other words, two people will come up with two different rankings for the possible alternatives and will have to find some way to resolve their differences, whereas an individual does not have this problem.

Browning and Chiappori support their argument with data from the Canadian Family Expenditure Survey (FAMEX). Overall, the survey does not support the Slutsky relations. But separating out single females and single males, each batch verifies the Slutsky relations.

The authors then ask themselves about two-person households. As mentioned before, any consumption decision will be the result of bargaining within the household, depending on the relative strengths and weaknesses of each partner (who brings in the money, and who has the strongest character, for example). What can be said about the outcome of such a process? Well, it should be efficient, which means that nothing will be thrown away — we cannot predict how the cake will be split, but we expect that everything will be eaten. This assumption is enough to write down a system of nonlinear partial differential equations which two-person household demand functions have to satisfy (see box). With the right mathematics, these relations can be shown to characterize com-

pletely the possible demand functions. And now the happy ending: the FAMEX data for two-person households support the Browning–Chiappori relations, in the same way that the FAMEX data for single people support the Slutsky relations. So the *Homo oeconomicus* model might well be true after all.

Some words about the right mathematics. The equations themselves are horrendous. The right way to handle them is to resort to a theory that was developed a hundred years ago for purely mathematical needs: exterior

differential calculus. At that time, mathematicians in France and Germany were trying to construct surfaces with certain geometrical properties, and they were led to very complicated systems of nonlinear partial differential equations. Appropriate methods to solve them were developed over fifty years, and finally codified by Cartan⁴. It is a beautiful piece of mathematics, which has been revived in an unexpected context.

The implications of Browning and Chiappori's work are strange. I can say, "Tell me what you eat, and I will tell you how many you are"; not "Tell me how much you eat, and I will tell you how many you are" — it would come as no surprise that many people eat more than do a few. What I am saying is that, if two households spend the same amount, I will be able to tell how many people there are in the household just by inspecting the consumption pattern. Even more remarkably, I will do this by checking whether a very complicated system of nonlinear partial differential equations is solved or not, using exterior differential calculus — another example of the versatility and power of mathematics. □
Ivar Ekeland is at the University of Paris-Dauphine, 75775 Paris, Cedex 16, France.

1. Antonelli, G. B. *Sulla Teoria Matematica delle Economia Politica* (Pisa, 1886); reprinted in *Giornale Degli Economisti e Annali di Economia* 10, 233–263 (1951).
2. Slutsky, E. *Giornale Degli Economisti* 51, 1–26 (1915).
3. Browning, M. & Chiappori, P. A. *Econometrica* (submitted).
4. Cartan, E. *Les Systèmes Différentiels Extérieurs et Leurs Applications Géométriques* (Hermann, Paris, 1945; reprinted 1971).

Membrane transport

Green light for Golgi traffic

Hugh R. B. Pelham

When eukaryotic cells secrete proteins, they do so by a circuitous route. The proteins are folded and assembled within a membrane-enclosed organelle, the endoplasmic reticulum (ER), whose membrane and internal contents are specialized for the job. They are then transported to the Golgi apparatus, and from there to the cell surface.

It is widely accepted that the intermediate steps in this process involve the budding and fusion of small vesicles. But the sites at which they form and fuse, and the number of such steps, have remained hotly debated topics. Now, however, in an elegant use of green fluorescent protein (GFP), Presley *et al.*¹ (page 81 of this issue) have answered a major outstanding question about the first step — from ER to Golgi. A few moments of time-lapse video* are enough to resolve an issue that years of microscopy on fixed cells have failed to settle. And the study also poses a new set of questions for the future.

Exit from the ER has long been a model

of vesicular transport². Electron microscopic images show, adjacent to the stacked cisternae of the Golgi complex, regions of the ER membrane that bud off to form vesicles (Fig. 1, overleaf). These vesicles then seem to fuse with a tubular network at the *cis* face of the Golgi apparatus, which is often referred to as the *cis*-Golgi network. Vesicles are formed from the ER by the action of the COPII (coat protein II) complex³, and they incorporate vesicle-targeting molecules (known as v-SNAREs). These proteins interact with corresponding t-SNAREs, which are present on the target membranes of the *cis*-Golgi network^{4,5}.

However, there is a twist. Proteins do not leave the ER only near to Golgi membranes — they emerge at sites that are distributed apparently at random, all over the cell. These sites are marked by little groups of vesicles and tubules^{2,6} (vesicular–tubular clusters, or VTCs), which have been extensively studied by electron microscopy. The relationship between these peripheral VTCs and the *cis*-Golgi network has been puzzling. They contain similar marker proteins, and both are sites at which escaped

*The video is at <http://dir.nichd.nih.gov/CBMB/pblabob.html>

resident proteins from the ER can be retrieved in vesicles that are coated with COPI proteins^{6,7}. But do secretory proteins pass through both the VTCs and the *cis*-Golgi network? If so, are there several vesicular steps, each with its own budding and targeting machinery? One view has been that the VTCs move along microtubules to the Golgi region, and join the tubular network there⁸. Another is that a second round of vesicle transport, mediated by COPI, carries cargo to the *cis*-Golgi network⁹.

It now turns out that the VTCs move. Presley *et al.*¹ fused GFP to the cytoplasmic tail of the ts045 mutant form of vesicular stomatitis virus glycoprotein, which is a favourite marker for transport studies. At high temperature this protein fails to fold correctly and it remains in the ER. On shift down to 32 °C, however, it rapidly exits, and a wave of traffic through the secretory pathway can be followed.

Luckily, the addition of GFP has very little effect on the transport of the viral glycoprotein, so the authors could directly watch the secretory process. They found that individual dots corresponding to peripheral VTCs formed, then moved along microtubules to the Golgi apparatus where they seemed to fuse into larger structures (Fig. 2). The moving elements were elongated and clearly larger than vesicles. Most surprising, perhaps, was the observation that once a VTC had left an exit point on the ER, that particular point was seldom — if ever — used again. Instead, new structures formed spontaneously, at apparently random locations.

Why is this result important? Not only does it dramatically show the role of the peripheral VTCs, but it also raises some fascinating questions. For example, what triggers the burst of active budding that forms the elements, and why does it cease equally suddenly? How can a structure that is derived entirely from the ER behave so differently just

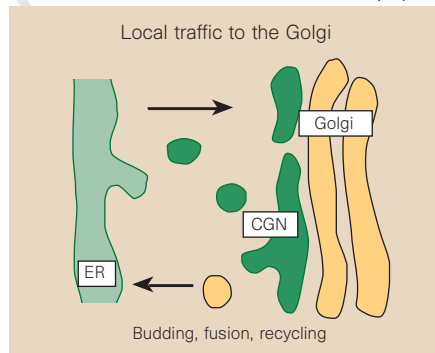


Figure 1 In an elegant study using green fluorescent protein, Presley *et al.*¹ have been able to observe directly the transport of proteins between the endoplasmic reticulum (ER) and the tubular network at the *cis* face of the Golgi apparatus (the *cis*-Golgi network, CGN). Over short distances, secreted proteins bud off in vesicles, which dock and fuse with the *cis*-Golgi network.

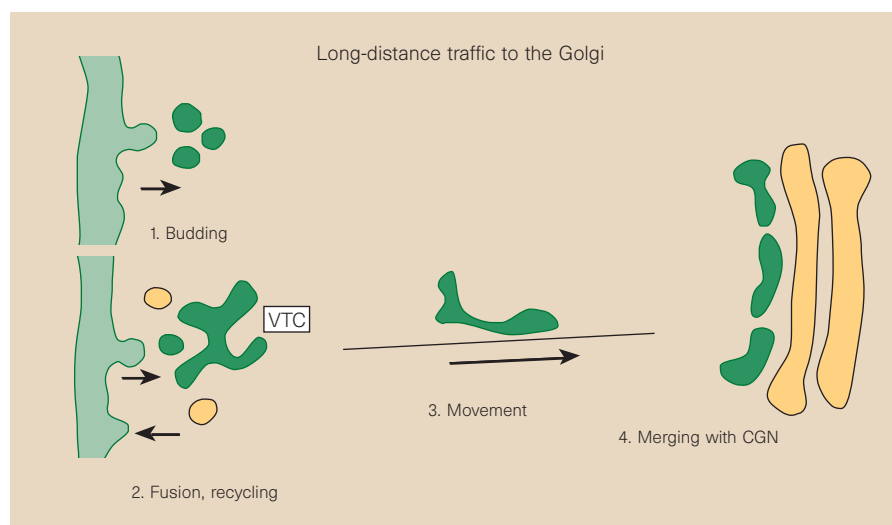


Figure 2 Proteins can also be transported from the ER to the Golgi over long distances, from sites that are randomly distributed around the cell. These sites are marked by groups of vesicles and tubules called vesicular–tubular clusters (VTCs). These VTCs, containing the secreted proteins, travel along microtubules to the Golgi apparatus, where they fuse with the *cis*-Golgi network. This is the first time that such transport has been directly observed, and the new study resolves the debate as to the function of the VTCs.

moments later — sorting out ER proteins, budding COPI vesicles and moving along microtubules in the opposite direction to ER membranes? Protein sorting during the budding step must ultimately be responsible for this change in character which may, perhaps, be initiated by the separation of ion pumps from counterbalancing channels, or of some lipid-modifying enzymes from others.

If these thoughts sound familiar, it is because the dynamics of the ER-to-Golgi step are strikingly similar to one view of endocytosis. Here, endocytic vesicles fuse to form early endosomes. Some proteins are then recycled to the cell surface while the remaining structures move to the Golgi region (for discussion see ref. 10). Similar mechanistic problems exist in each case.

A further implication of the work by Presley *et al.*¹ is that the v- and t-SNAREs, which are thought to mark vesicles and their targets⁴, respectively, have less distinct functions than has been assumed. The *de novo* formation of VTCs implies that ER-derived vesicles fuse with one another. Fusion is likely to be driven by t-SNARE/v-SNARE interactions and, indeed, the 'early-Golgi' t-SNARE syntaxin 5 is found in VTCs (J. Stinchcombe and C. Hopkins, personal communication), which it must reach by recycling through the ER. So the initial fusion step in the secretory pathway (as in the endocytic pathway) may be between equivalent membranes that bear both t- and v-SNAREs. There is a precedent for this event in the homotypic fusion of yeast vacuoles¹¹.

The spontaneous generation of membranes resembling the *cis*-Golgi from the ER is, for those who have seen the movie, a startling and unforgettable sight. It

bears on another controversy, concerning the mechanism by which proteins pass through the stacked cisternae of the Golgi complex^{6,12}. One model, derived from electron microscopic studies, invokes cisternal maturation. This requires the continuous generation of new, cargo-carrying cisternae which, as they pass through the Golgi stack, are modified by vesicle-mediated delivery and removal of components. They will eventually be converted into post-Golgi carriers.

The events seen by Presley *et al.*¹ could plausibly be interpreted as being the first step in this process. However, they do not rule out the alternative possibility — that cisternae have a stable existence, and that cargo is carried between them in vesicles. The delivery of VTCs to the *cis*-Golgi network would then do no more than compensate for the loss of membrane in vesicular form. Unfortunately, light microscopy cannot achieve a sufficiently high resolution to allow a distinction between vesicular and cisternal transport within the Golgi stack to be made, and the debate will continue. Spectacular though it is, GFP technology has its limits. □

Hugh R. B. Pelham is in the MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

1. Presley, J. E. *et al.* *Nature* **389**, 81–85 (1997).
2. Jamieson, J. D. & Palade, G. E. *J. Cell Biol.* **34**, 577–596 (1967).
3. Barlowe, C. *et al.* *Cell* **77**, 895–907 (1994).
4. Rothman, J. E. *Nature* **372**, 55–63 (1994).
5. Banfield, D. K. *et al.* *J. Cell Biol.* **127**, 357–371 (1994).
6. Bannykh, S. I. & Balch, W. E. *J. Cell Biol.* **138**, 1–4 (1997).
7. Stinchcombe, J. C. *et al.* *J. Cell Biol.* **131**, 1387–1401 (1995).
8. Saraste, J. & Svensson, K. *J. Cell Sci.* **100**, 415–430 (1991).
9. Kreis, T. E., Lowe, M. & Pepperkok, R. *Annu. Rev. Cell Dev. Biol.* **11**, 677–706 (1995).
10. Gruenberg, J. & Maxfield, F. R. *Curr. Opin. Cell Biol.* **7**, 552–563 (1995).
11. Nichols, B. J. *et al.* *Nature* **387**, 199–202 (1997).
12. Schekman, R. & Mellman, I. *Cell* **90**, 197–200 (1997).



100 YEARS AGO

Among many other papers in the *Proceedings of the Indiana Academy of Science*, dated 1894, but only just received, is one by Mr. D. T. MacDougal, showing that various species of *Cypripedium* have an irritant action upon the human skin. It was found that when the leaves of *C. spectabile* were rubbed lightly upon the skin of the wrist, arm, face, or ear, the person experimented upon was usually "poisoned" in a degree corresponding to the manner of application, and in a time varying from ten to twelve hours. There could be no doubt about the unpleasant effects produced by the leaves, for Mr. MacDougal soon found that he could not obtain subjects willing to sacrifice their feelings upon the altar of scientific knowledge. ... To ascertain whether the effect was due to the mechanical injury resulting from piercing the skin by the pointed hairs upon the leaves, or to the corrosive action of the secretion found on the outside of the globular tips of the glandular hairs ... the hairs of each kind were taken from the leaf by means of a fine pair of forceps, and the tip pressed against the skin. Irritation was found to result from the contact of the glandular hair only.

From *Nature* 2 September 1897.

50 YEARS AGO

It is common knowledge that open coal fires radiate only a small fraction of the heat available from bituminous coal, most of the rest being carried up flues to the outside air. In many countries climatic conditions or the need for thrift enforce various means to avoid waste of fuel. Independent stoves are placed in living-rooms or heat is distributed throughout buildings by means of warm water or air propelled either mechanically or by gravitational action. To-day circumstances are compelling Britain increasingly to adopt similar methods. The growing use of closed stoves is a familiar example whereby 65 per cent of the heat of combustion can be distributed in a simple manner, at the cost of losing, wholly or partly, the heating by radiation from the glowing fuel. ... The widespread construction of new houses offers advantages in such methods, advantages which are being taken.

From *Nature* 6 September 1947.

Climate change

Icy message from the Antarctic

Eugene Murphy and John King

Historically, much of the whaling in the Antarctic occurred in the vicinity of the ice edge where the whales congregated to feed on krill. During the austral spring, the whalers tracked the animals as they moved south with the edge of the sea ice, and their records of harvesting position thus provide proxy data for ice extent before monitoring by satellite began. An analysis of those records by de la Mare, published on page 57 of this issue¹, suggests that there was a large and rapid decrease in Antarctic summertime sea-ice extent between the 1950s and 1970s. The cause of this inferred decrease is unknown. But it will be sadly ironic if the human slaughter of the great whale populations reveals something of the Earth's climate system that may have dramatic implications for our own long-term future on the planet.

Such a change in sea-ice extent would have global significance, because ice cover of the polar oceans is an important component of the Earth's climate system. It exerts strong controls on the exchanges of energy between atmosphere and oceans at high latitudes. Sea ice has a higher albedo than the open ocean and thus modifies the energy balance of polar regions. It also acts as an insulating blanket, reducing the transfer of heat from the underlying oceans to the cold polar atmosphere.

These processes introduce a positive feedback into the climate system, making the climate naturally variable in polar regions and, potentially, making them particularly sensitive to changes in forcing such as those caused by increasing anthropogenic emissions of greenhouse gases. From the data available so far², and some modelling studies³, it seems that the Antarctic sea ice may be relatively insensitive to climate change compared with that in the Arctic, because of different potential physical feedbacks. But we know so little, and any evidence that can contribute to the debate is of great value.

Much of our thinking on the stability of Antarctic sea ice is based on the comprehensive observations of its extent that have been available only since 1973, when satellite monitoring began. That record² does not reveal any significant overall trend, but it does show a great deal of regional variability^{4,5}. Such behaviour makes it difficult to reconstruct overall Antarctic sea-ice coverage, and de la Mare has been careful to address this problem by including confounding effects in his statistical analyses.

There is little other information with which to compare his conclusions. Work⁶

published in 1981, based on the first few years of satellite data and earlier ship-based records, suggested that there was only a modest decline in summer sea-ice extent. Regional studies can give a very different view. Earlier this year, for example, a detailed analysis⁷ of the sea ice in the Bellingshausen Sea area to the west of the Antarctic Peninsula highlighted the coupled nature of the ocean-ice-atmosphere system generating the variation, and found that there has been a reduction in sea-ice extent in the past couple of decades.

One of the expected consequences of a summer sea-ice retreat would be an increase in atmospheric and oceanic temperatures at high latitudes, yet a study⁸ of Southern Ocean sea-surface temperatures measured by Japanese whaling ships failed to detect any significant warming in the 60 °S–70 °S latitude band over the period 1946–84. However, if the whalers followed the ice edge, and hence remained in waters of a similar temperature, then the observation is not inconsistent with a southward shift in sea ice.

Of other approaches, climate-model experiments^{9,10} involving reductions in sea-ice extent or percentage cover indicate the consequences of small warmings and increased precipitation over continental Antarctica, along with increasing wind stress which could affect the strength of the Antarctic Circumpolar Current and other aspects of the ocean circulation. The relevant data are again limited. Temperature records from most Antarctic stations show small (and generally statistically insignificant) warming trends¹¹; and the growing evidence¹² that there have been increases in precipitation in parts of Antarctica over the past few decades cannot yet tell us much about the larger-scale and longer-term picture.

Sea ice might have a further effect on ocean circulation through its possible role in the formation of so-called bottom water. As sea water freezes, brine-rich water is rejected and sinks to form a dense water mass over the continental shelf. Subsequent mixing of this water with surrounding water masses produces Antarctic Bottom Water (AABW), which is a major component of the global deep-ocean circulation. The chain of processes leading to bottom-water formation is complex, but it is likely that a reduction in the rate of winter sea-ice production would affect it.

In this context, a 1996 analysis¹³ of the characteristics of the AABW entering the Argentine Basin has indicated that there are changes between years, which are penetrating to lower latitudes and may therefore



Figure 1 Has the whales' tale got a sting in it — that of a large and rapid decrease in Antarctic summer sea-ice coverage in the mid-twentieth century?

affect global climate. These effects have been explained by changes in the circulation pattern in the Weddell Gyre where much of the AABW is formed. The improvement of sea-ice models, and the representation of processes occurring in the Southern Ocean in general, remain a priority for climate modellers.

The main value of de la Mare's work is in further highlighting the issue of the mechanisms involved in generating variation and potentially rapid change in the Southern Ocean. It is ten years since Broecker¹⁴ pointed out that we should expect "Unpleasant surprises in the greenhouse" as a result of rapid changes in climate systems. There is now increasing evidence globally supporting the view that such rapid changes in the Earth's climate systems can occur naturally, and indeed such changes have probably taken place in the past in the Southern Ocean¹⁵.

This evidence indicates that the variability inferred by de la Mare may be natural and not connected to any human-induced changes. But as yet we do not know. With the characterization^{4,5} of the spatial and temporal coherence of the interannual and sub-decadal cyclicity of the sea ice, ocean and atmosphere in the Southern Ocean, variability in physical and ecological systems has rightly become the subject of a great deal of research attention. In particular, the question of whether there are several steady states between which these systems can jump needs to be addressed sooner rather than later.

Is there a message in all this for the management of large ecological systems in general, and the whales of the Southern Ocean in particular? Perhaps, but we may find it unpalatable. The decline in Antarctic sea ice, inferred by de la Mare, occurred at the time of the final reduction in whale numbers. There could therefore have been irreversible shifts in this system and the recovery of whale

populations to pre-exploitation levels may not now be possible. There may well have been dramatic changes in the rest of the ecosystem, for instance in the pattern of primary production, krill recruitment and the distribution and abundance of higher predators. We might just have to accept that rapid

Epidemiology

A human germ project?

John Danesh, Robert Newton and Valerie Beral

Infectious agents that lead to chronic diseases or cancer tend to persist silently for many years before causing disease in a small proportion of their hosts. There is now much interest in the possible link between coronary heart disease and one such infectious agent, *Chlamydia pneumoniae*, an intracellular bacterium. Studies published in the *Lancet*¹ and in *Circulation*² report the first randomized trials of antibiotics in the prevention of coronary heart disease. Their findings — that antibiotics with activity against *C. pneumoniae* may prevent further heart attacks in people suffering from heart disease — were based on only a small number of patients, so they are prone to chance fluctuations. But these studies have prompted the start of large-scale clinical trials, and they highlight a larger trend in clinical research, towards the identification of new associations between common infectious agents and chronic disease (Table 1, overleaf).

Interest in infections as potential causes of human cancer and other chronic disease has waxed and waned for almost a century, since F. P. Rous demonstrated the carcinogenic effects of chicken sarcoma virus. The search for human 'cancer viruses' intensified in the 1960s, culminating in President

changes can occur, making the management of marine systems a short-term task in which we must be prepared to manage change rather than expecting to maintain a status quo. Setting out the criteria for management in such circumstances will be an interdisciplinary task, and a challenge for both scientists and politicians. □

Eugene Murphy is in the Marine Life Sciences Division, and John King in the Ice and Climate Division, British Antarctic Survey, NERC, Madingley Road, Cambridge CB3 0ET, UK.

1. de la Mare, W. K. *Nature* **389**, 57–60 (1997).
2. Björge, E., Johannessen, O. M. & Miles, M. W. *Geophys. Res. Lett.* **24**, 413–416 (1997).
3. Stouffer, R. J., Manabe, S. & Bryan, K. *Nature* **342**, 660–662 (1989).
4. Murphy, E. J., Clarke, A., Symon, C. & Priddle, J. *Deep-Sea Res.* **42**, 1045–1062 (1995).
5. White, W. B. & Peterson, R. G. *Nature* **380**, 699–702 (1996).
6. Kukla, G. & Gavin, J. *Science* **214**, 497–503 (1981).
7. Jacobs, S. S. & Comiso, J. C. *J. Clim.* **10**, 697–709 (1997).
8. Mierzejewska, A. W., Wu, Z., Newell, R. E. & Miyahita, T. *Bull. Am. Meteorol. Soc.* **78**, 443–447 (1997).
9. Mitchell, J. F. B. & Senior, C. A. Q. *J. R. Meteorol. Soc.* **115**, 225–246 (1989).
10. Simmonds, I. & Budd, W. F. Q. *J. R. Meteorol. Soc.* **117**, 1003–1024 (1991).
11. Jones, P. D. *Geophys. Res. Lett.* **22**, 1345–1348 (1995).
12. Morgan, V. I., Goodwin, I. D., Etheridge, D. M. & Wooley, C. W. *Nature* **354**, 58–60 (1991).
13. Coles, V. J., McCartney, M. S., Olson, B. B. & Smethie, W. M. J. *Geophys. Res.* **101**, 8957–8970 (1996).
14. Broecker, W. S. *Nature* **328**, 123–126 (1987).
15. Barker, P. F. *Phil. Trans. R. Soc. Lond. B* **338**, 259–267 (1992).

Nixon's National Cancer Act in 1971. But only limited success followed and, by the early 1980s, interest began to fade. In retrospect, progress may have been hindered by technical limitations. Now, molecular technologies such as the polymerase chain reaction (PCR) can both identify previously unrecognized infections and indicate the presence of recognized organisms in unexpected tissues. And epidemiological methods, such as large observational studies and clinical trials, provide a way of distinguishing between causal and coincidental associations.

Microbiologists have traditionally tried to isolate and then multiply organisms in culture. But only a few per cent of all bacteria and viruses can be cultured by standard methods³, so this approach has revealed only a small sample of microbial life. Instead, improved microscopy has opened new possibilities: the Epstein-Barr virus (Table 1) was discovered in 1964 when it was directly seen by electron microscopy (EM) in cells from Burkitt's lymphoma; EM studies indicated the presence of curved bacteria in stomach mucosa nearly a decade before the viability of *Helicobacter pylori* was established by culture⁴; and herpesviruses and *C. pneumoniae* were first detect-

Solar physics

Down the sunspot plughole

An 88-year-old mystery surrounding sunspots has finally been solved, thanks to the ingenious application of an optical tomographic imaging technique.

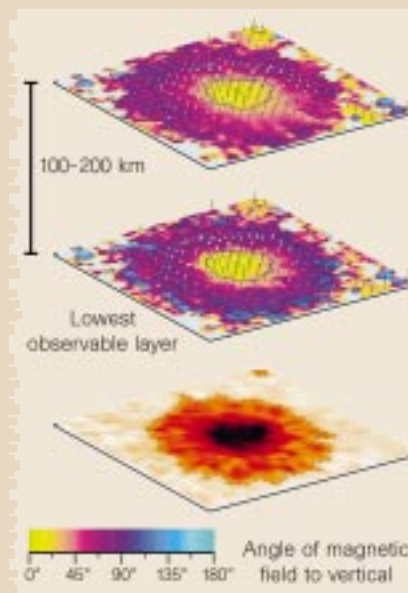
Sunspots are relatively cool regions of the solar disk where bundles of magnetic field lines break through the surface. A puzzling phenomenon, first noted by J. Evershed in 1909 (*Mon. Not. R. Astron. Soc.* 69, 454–457; 1909), concerns the apparent violation of one of the basic laws of physics — conservation of mass. Spectroscopic observations of sunspots show a flow of material streaming across the face of the spot, only to perform a mysterious vanishing act at its outer edge.

But new false-colour maps (right) of a typical sunspot, obtained by C. Westendorp Plaza *et al.* and published in this issue (*Nature* 389, 47–49; 1997), reveal the secret of these solar Bermuda Triangles. The images were constructed using an indirect tomographic method, which involves mathematical processing of optical polarization data. The results, at different depths in the solar atmosphere, are shown in the upper two panels. (The bottom panel is a conventional optical image of the spot.) Importantly, the technique yields information not only

about the velocity of material within the vicinity of a sunspot, but also about the magnetic field (represented by arrows).

The new tomography implies that the field lines form closed loops in the deepest layers of the atmosphere, near to the solar surface. This is a crucial observation, because theory suggests that gas should be physically routed along these field lines, or 'flux tubes'. And indeed, in the outermost regions of the spot, inflowing material is strongly correlated with inwardly directed magnetic flux (the blue-shaded regions). So the sink of material is at last revealed — at the edge of the spot, the low-lying magnetic flux tubes drag the gas back into the Sun, rendering the flow spectroscopically invisible.

Controversy has often surrounded the Evershed effect — for example, some earlier observations seemed to contradict it, in that they failed to indicate an abrupt termination of the flow at the spot edge. But now the reason is clear. The key point is that physical quantities (such as magnetic flux) vary with height within the solar atmosphere. Previous measurements were incomplete in that they could only sample one particular layer, as in the fabled story in which an elephant is described by blind people. But the 3D tomographic technique can probe these gradients,



and measure the field and mass flow within the deepest atmospheric layers.

It is now clear from the tomography that, in the higher atmospheric layers, the magnetic field lines are open, rather than being closed loops, and extend beyond the visible limits of the spot. Therefore, at this height, the flow should not abruptly terminate at the spot edge, which explains the apparently anomalous earlier observations of these layers.

So in more ways than one, the Evershed flow could be said to be an open-and-shut case.

Karen Southwell

ed in the atherosclerotic plaques of blood vessels by EM⁵.

The various subtypes of human papillomavirus were discovered only with the advent of DNA molecular hybridization and cloning in the early 1980s. More recently, an elegant variation of PCR was used to identify human herpesvirus-8 in the lesions of patients with Kaposi's sarcoma⁶. PCR studies have also identified the bacteria that are thought to cause Whipple disease, and those that are responsible for cat-scratch disease⁷. Other variations, such as 'kingdom'-specific PCR, are also being used in attempts to detect pathogens in various tissues. This technique involves probes

against the highly conserved bacterial 16S ribosomal RNA gene, and it is being used to hunt for unrecognized or unexpected bacteria in joints and arterial walls⁸.

In the latest study of *C. pneumoniae* and the heart, Gurfinkel *et al.*¹ randomized people suffering from heart disease — irrespective of whether they had high serum concentrations of antibody to the infectious agent — to either placebo or a short course of the antibiotic roxithromycin. Such a blanket strategy of treatment could be useful in larger trials, particularly if it involved antibiotics that could also eliminate other persistent infectious agents (such as *H. pylori*) which might be associat-

ed with heart attacks. Also, roxithromycin might interfere with the host's inflammatory responses, which could be important in mediating any effect of *C. pneumoniae* on blood vessels. However, of the few hundred people involved in the studies of Gurfinkel *et al.*¹ and Gupta *et al.*², fewer than two dozen developed heart attacks or cardiac complications, so the samples were too small to be informative.

The mechanisms by which *C. pneumoniae* might damage the heart remain poorly understood, but it is not unusual for mechanistic evidence to lag behind epidemiological findings in studies of chronic disease. For example, some of the molecular pathways by which cigarette smoking causes lung cancer have been worked out only very recently — decades after the causal link with disease was established. And there is still uncertainty about how *H. pylori* causes peptic ulceration, or how human herpesvirus-8 causes Kaposi's sarcoma.

The presence of an infectious agent in diseased tissue — such as markers of *C. pneumoniae* in the walls of blood vessels — does not necessarily imply a causative role. So when do correlations justify the inference of causation? Classical microbiologists

Table 1 Examples of common persistent infections with associated chronic diseases

Date	Infectious agent	Associated disease
1964	Epstein-Barr virus	Burkitt's lymphoma, nasopharyngeal carcinoma
1965	Hepatitis B virus	Hepatitis, liver cancer
1983	Human immunodeficiency virus-1	AIDS
1983	<i>Helicobacter pylori</i>	Chronic gastritis, peptic ulceration, gastric carcinoma?
1983	Human papillomavirus (mainly types 16 and 18)	Cervical, penile, vulvar and anal cancer
1983	<i>Borrelia burgdorferi</i>	Lyme disease
1986	<i>Chlamydia pneumoniae</i>	Respiratory disease, coronary heart disease? asthma?
1989	Hepatitis C virus	Hepatitis, liver cancer
1994	Human herpesvirus-8	Kaposi's sarcoma



Figure 1 Coloured transmission electron micrograph of hepatitis B virus particles. The orange spheres are the complete, active virus; brown rods are part of the disassembled virus coat, without a core, and they are inactive.

invoke Koch's postulates, but these are often unhelpful in chronic diseases because most causal factors are neither necessary nor sufficient for those diseases to occur. Moreover, experimental inoculation of animals is often impossible, or it is irrelevant to human disease. For example, the link between *H. pylori* and peptic ulceration was demonstrated by randomized clinical trials, and the human immunodeficiency virus was shown to cause AIDS by observational epidemiology, rather than by fulfilment of Koch's postulates.

Trials of anti-infective treatments are useful for testing hypotheses about the roles of infection in disease, and their findings have led to the routine use of antibiotics in the treatment of peptic ulceration, Lyme arthritis and, in some cases, even B-cell gastric lymphomas. Vaccines and antibiotics are also being tested in studies for the prevention of liver cancer (hepatitis B vaccine)⁹, gastric cancer (*H. pylori* eradication regimens)¹⁰ and cervical cancer (human papillomavirus vaccines)¹¹.

When the development of an infection may not be preventable, or its effects cannot be fully or rapidly reversed by such treatments, observational epidemiology can be useful in predicting the likelihood of causality. The numerical strength of epidemiological associations provides a particularly good indicator — for example, lung cancer is 20 times more common in cigarette smokers than in non-smokers. Yet, for some persistent infections, epidemiological studies have reported greater than

100-fold associations, such as for hepatitis B virus (Fig. 1) and liver cancer¹², or human herpesvirus-8 and Kaposi's sarcoma¹³. Moreover, human papillomavirus types 16 or 18 are about 30 times more common in women with cervical cancer¹⁴, and *C. pneumoniae* is around ten times more common in atherosclerotic plaques than in disease-free blood vessels.

What diseases might next be linked to persistent infections? Because most chronic diseases tend to have long induction periods, and most persistent infections behave insidiously, few clinical or epidemiological features can be used as definite guides. But certain features of disease may offer clues. These include a high incidence of disease in people who are prone to infections (such as the immunosuppressed); large variations in incidence by geographical region or other patterns of clustering; apparent improvement with antimicrobial treatments; and strong correlations with markers of poverty. On the basis of these and other features, disease that have been suspected of being (at least partly) caused by infectious agents include Crohn's disease, rheumatoid arthritis, multiple sclerosis, chronic obstructive lung diseases, multiple myeloma, certain leukaemias and lymphomas, and melanoma^{7,15–17}.

All of these discoveries make a case for a coordinated effort by microbiologists, virologists, molecular biologists, clinicians and epidemiologists to search for new — as well as known — infectious agents in human tissues, and to study their possible correlations with clinical disease. Many human pathogens have the potential to cause chronic diseases, and new technology, combined with modern epidemiology, has made the discovery of new associations more feasible than ever before. □

John Danesh is in the Clinical Trial Service Unit and Epidemiological Studies Unit, University of Oxford, Oxford OX2 6HE, UK.

Robert Newton and Valerie Beral are in the Imperial Cancer Research Fund Cancer Epidemiology Unit, Oxford OX2 6HE, UK.

- Gurfinkel, E., Bozovich, G., Daroca, A., Beck, A. & Mautner, B. *Lancet* **350**, 404–407 (1997).
- Gupta, S. et al. *Circulation* **96**, 404–407 (1997).
- Service, R. F. *Science* **275**, 1740–1742 (1997).
- Goodwin, C. S., Mendall, M. & Northfield, T. C. *Lancet* **349**, 265–269 (1997).
- Melnick, J. L. et al. *Lancet* **ii**, 644–647 (1983).
- Chang, Y. et al. *Science* **266**, 1865–1869 (1994).
- Lorber, B. *Ann. Intern. Med.* **125**, 844–851 (1996).
- Kingsley, G. *Lancet* **349**, 1038–1039 (1997).
- Chang, M. H. et al. *N. Engl. J. Med.* **336**, 1855–1859 (1997).
- Danesh, J., Forman, D., Collins, R. & Peto, R. *Lancet* **348**, 758–759 (1996).
- McNeil, C. J. *Natl Cancer Inst.* **89**, 281–282 (1997).
- Beasley, R. P., Hwang, L. Y., Lin, C. C. & Chien, C. S. *Lancet* **ii**, 1129–1133 (1981).
- IARC Working Group 70–87 (1996).
- Munoz, N. et al. *Int. J. Cancer* **52**, 743–749 (1992).
- Krause, A., Kamradt, T. & Burmester, G. R. *Curr. Opin. Rheum.* **8**, 203–209 (1996).
- Rettig, M. B. et al. *Science* **276**, 1851–1854 (1997).
- Kinlen, L. J. *Br. J. Cancer* **71**, 1–5 (1995).

Daedalus

No-party democracy

A democratic election, said J. B. S. Haldane, is a way of deciding the result of a civil war without having one. He had a point. In some developing countries political parties are essentially tribal groupings, and an election is a formalized tribal fight. Even in the developed world, the parties used to stand for identifiable social blocs — the workers, the middle class, and so on — and each party favoured its own. This at least gave the voters a choice. But modern parties, as shown in the recent British election, have quite a different strategy. They 'steal each other's clothes'. Each tries to broaden its appeal by devising policies very close to those of its rivals. The result is a complex 'curd' of policies, intertwined and very similar — indeed, often identical. This is the notorious 'Maas-trick', in which all parties collude in the same policy, denying voters any way of rejecting it. But such is the logic of modern consensual politics; so Daedalus is now taking that trick to extremes.

Group decisions are best reached, he says, not by simple voting, but by the Delphi method used in technological forecasting. A secret poll discovers the spread of opinion, which is then displayed to all the voters. Those finding themselves on the extremes of the distribution tend to change their minds; those in the mainstream retain their convictions. So a second round of voting gives a narrower distribution, which is displayed again for a third round of voting. A few iterations usually produce a clear consensus.

Daedalus wants to do this for a whole nation, and the whole range of political policies. Modern media, especially the Internet, seem ideal for the purpose. With DREADCO's computing expertise behind it, a consortium of polling firms could probably do the job whenever an election loomed. All parties could then be presented with one comprehensive set of policies expressing the people's current will. They could only compete by submitting bids to carry them out for the least amount of tax.

Politicians will heave a sigh of relief at not having to invent any more damned policies. They will compete purely for power, which is all they want anyway, and will be judged on their cash flow and competence, which is all that matters in practice. But as the logic of the customer-contractor relationship sinks in, they could find themselves facing novel competition. Bill Gates or Richard Branson might put in a lower bid than any of them.

David Jones

Odour conveys status on cockroaches

The evolution of social dominance is thought to require phenotypic traits that convey honest information regarding an individual's status¹. These 'badges of status' could be communicated by any means, but only avian plumage has been examined and provides equivocal support for the hypothesis². By chemical manipulation of the odour of socially naive male cockroaches (*Nauphoeta cinerea*) we show that the sex pheromone of males determines rather than reflects status. The individual chemical constituents of the pheromone have contrasting effects on status, but the evolution of the blend that confers dominance may be constrained by contrasting selection pressures resulting from female mate choice.

The male-produced pheromone of *N. cinerea* that attracts females is composed of three compounds: 3-hydroxy-2-butanone, 2-methylthiazolidine and 4-ethyl-2-methoxyphenol^{3,4}. The quantities of these are heritable and co-vary with male status^{5,6}. Female mate choice is based on pheromonal differences among males^{7,8}. We manipulated the quantities of these three components

individually and in combinations to investigate their effect on interactions between males.

Adult males were isolated on emergence, and a filter-paper circle (6 mm diameter) was glued to the pronotum. The pheromonal profile of males was manipulated by adding individual pheromone components: a combination of the two developmentally and genetically correlated components^{5,6}, a blend of all three components, or acetone (control) onto the filter paper. The quantity applied was based on the mean quantity of pheromone isolated from the sternal pheromone glands of dominant males^{5,6}. After a 15-min recovery period, a control male and a manipulated male were placed in an open arena (17 × 12 × 6.5 cm) for observations. The social status of each male was determined on the basis of patterns of agonistic behaviour⁵⁻⁹.

Status depended on the specific pheromone manipulation (Fig. 1). When 3-hydroxy-2-butanone alone was increased in quantity, males were likely to be treated and to react as subordinates. When either 2-methylthiazolidine or 4-ethyl-2-methoxyphenol was increased, there was an increased likelihood of being dominant. This effect was even greater when the quantities of both compounds were increased together. The influence of 3-hydroxy-2-butanone is cancelled by the combined effects of 2-methylthiazolidine and 4-ethyl-2-methoxyphenol; increasing all three compounds simultaneously did not affect status. Thus, the compounds have individual, additive and contrasting effects on status.

Our manipulations did not result in abnormal behaviour. The direction and intensity of interactions were stable within a one-minute observation period. Manipulated males behaved the same as control males of the same status and there were no obvious differences in the way that males were treated. There were no significant differences in the frequency of aggressive acts by males scored as dominant or the frequency of submissive acts by subordinate males. Highly aggressive acts associated with uncertain initial status, such as grappling, biting and kicking, were no more frequent in those interactions resulting in the manipulated male being scored as dominant as in the interactions where the control male was scored as dominant.

A dual function allows badges of status to be 'honest' signals^{1,10}. The sex pheromone of *N. cinerea* may have a dual function, acting in both male-male competition and female

mate choice⁵⁻⁹. Male-male competition influences mating success in *N. cinerea*^{5,9} so selection could be expected to increase the quantities of both 2-methylthiazolidine and 4-ethyl-2-methoxyphenol and decrease the quantity of 3-hydroxy-2-butanone. There is no genetic constraint to this pattern of evolution: the quantities of 2-methylthiazolidine and 4-ethyl-2-methoxyphenol are genetically correlated but the quantity of 3-hydroxy-2-butanone is genetically independent of these and is presumably free to evolve⁶. However, female *N. cinerea* also use all three components of the sex pheromone to discriminate among males^{5,6,10}. Sex differences in either the signal perception or the functions of the signal components can thus maintain variations in this pheromone and its multicomponent structure.

Patricia J. Moore

Division of Natural Sciences and Mathematics, Transylvania University, Lexington, Kentucky 40508, USA

Nancy L. Reagan-Wallin

Division of Math and Sciences, Kentucky State University, Frankfort, Kentucky 40601, USA

Kenneth F. Haynes, Allen J. Moore

Department of Entomology, University of Kentucky, Lexington, Kentucky 40546-0091, USA
e-mail: ajmoore@pop.uky.edu

1. Dawkins, R. & Krebs, J. R. in *Behavioural Ecology* (eds Krebs, J. R. & Davies, N. B.) 282-309 (Sinauer, Sunderland, MA, 1978).
2. Whitfield, D. P. *Trends Ecol. Evol.* **2**, 13-18 (1987).
3. Sreng, L. *J. Chem. Ecol.* **16**, 2899-2912 (1990).
4. Sirugue, D., Bonnard, O., LeQuere, J.-L., Farine, J.-P. & Brossut, R. *J. Chem. Ecol.* **18**, 2261-2276 (1992).
5. Moore, A. J., Reagan, N. L. & Haynes, K. F. *Anim. Behav.* **50**, 191-202 (1995).
6. Moore, A. J. *Evolution* (in the press).
7. Moore, A. J. *Anim. Behav.* **36**, 303-305 (1988).
8. Moore, A. J. & Moore, P. J. *Evolution* **42**, 387-391 (1988).
9. Schal, C. & Bell, W. J. *Biol. Behav.* **8**, 117-139 (1983).
10. Johnstone, R. A. & Norris, K. *Behav. Ecol. Sociobiol.* **32**, 127-134 (1993).

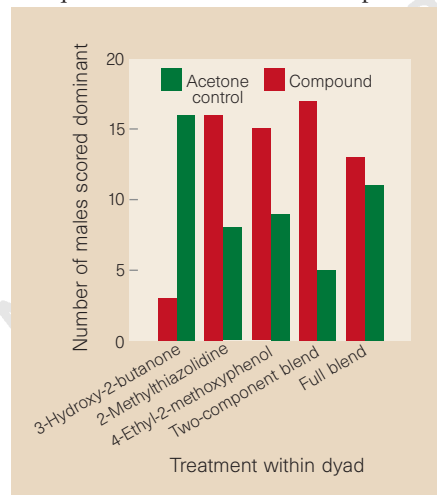


Figure 1 Pairs of males were treated, one with the compound (or mixture) and one with acetone (10 µl placed on a filter paper on the pronotum). The experiment was performed blind, with respect to which chemicals were being applied. Males with increased 3-hydroxy-2-butanone were more likely to be treated and act as subordinate (binomial test, $P=0.005$). In contrast, males with increased 2-methylthiazolidine ($P=0.054$) or 4-ethyl-2-methoxyphenol ($P=0.115$) were slightly more likely to be treated as and to act dominant. This effect was stronger for males treated with increased 2-methylthiazolidine and 4-ethyl-2-methoxyphenol ($P=0.005$). Increasing the amount of all three compounds in ratios corresponding to the individual manipulations had no effect on status ($P=0.345$).

Anandamide may mediate sleep induction

We report here that oleamide, a putative sleep factor¹, and anandamide, an endogenous cannabinoid ligand^{2,3}, cause similar pharmacological effects in mice. Only anandamide binds to the cannabinoid CB₁ receptor but by inhibiting the enzyme which inactivates anandamide (thus increasing its concentration), oleamide potentiates anandamide binding to CB₁ enhancing anandamide effects in mice. Our observations raise the possibility that some oleamide effects (including induction of sleep) may be mediated by anandamide.

Cravatt *et al.* identified and reported the sleep-inducing properties of oleamide, a lipid found in the cerebrospinal fluid of

sleep-deprived cats¹; and we have isolated and characterized anandamide, a lipid from the brain which is an agonist for the cannabinoid receptor^{2,3}. Drowsiness or sleepiness are well-known effects in the later stages of intoxication by marijuana, whose active constituent, Δ^9 -tetrahydrocannabinol (Δ^9 -THC), and anandamide have a close biochemical and behavioural profile. Both substances bind to the brain cannabinoid receptor, CB₁ (refs 2–4).

The sleep-producing properties of anandamide are not known, but Santucci *et al.*⁵ have found that the CB₁ antagonist SR141716A increases the time spent awake at the expense of both slow-wave and rapid-eye-movement sleep. They suggested that "...an endogenous cannabimimetic (anandamidergic?) system may regulate the organization of the sleep–waking cycle".

To establish whether there is any relationship between anandamide and oleamide we compared their *in vivo* and *in vitro* effects. We examined oleamide in four assays commonly used together for testing cannabinoid (including anandamide) activity^{3,6}: the ring immobility (catalepsy) test, which measures the percentage of time mice remain motionless; the open-field test, which measures locomotor activity; hypothermia; and response to a hot plate (antinociception). The median effective doses (ED₅₀ values) obtained from these tests (Table 1) show that oleamide has essentially the same activity profile as anandamide, though it is less potent in most tests. Testing anandamide in the presence of 7.5 mg oleamide per kg body mass (Table 1) showed that oleamide potentially increased anandamide activity. But oleamide is not a cannabinoid, as even at 10 μ M, it did not bind to CB₁, confirming a previous report⁷.

The enzyme responsible for anandamide inactivation, fatty-acid amide hydrolase (FAAH), recognizes oleamide as a substrate⁸, and its molecular characterization, cloning and expression have been reported recently⁹. Using the assays described in refs 8 and 10, we found that oleamide inhibited FAAH-mediated hydrolysis of [¹⁴C]anandamide by mouse neuroblastoma N₁₈TG₂ cells in a dose-dependent manner (Fig. 1a). With the 10,000g particulate fraction, which contains most of the cell FAAH⁸, the effect was significant at an oleamide concentra-

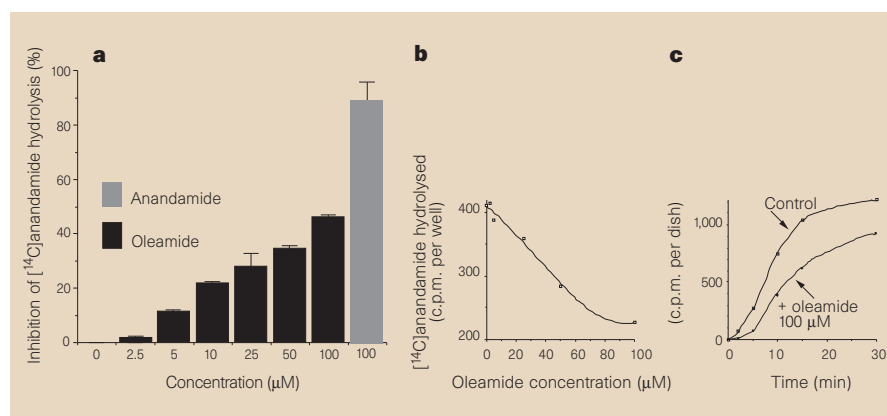


Figure 1 **a**, Dose-dependent inhibition of [¹⁴C]anandamide hydrolysis by N18TG2 cell 10,000g particulate fractions. **b**, Dose-dependent inhibition of [¹⁴C]anandamide hydrolysis by intact confluent N18TG2 cells. **c**, Effect of 100 μ M oleamide on time-dependent [¹⁴C]anandamide hydrolysis by intact confluent N18TG2 cells. The breakdown of [¹⁴C]anandamide was measured from the [¹⁴C]ethanolamine released by hydrolysis of the amide bond in anandamide as described previously⁹. Data in **a** are means $\pm$ s.e.m. of three separate experiments. Data in **b** and **c** are means of duplicates, representative of three experiments.

tion of 5 μ M, with 48% inhibition with 100 μ M oleamide. This effect was smaller than that obtained with anandamide (90% inhibition with 100 μ M), in agreement with previous observations that FAAH catalyses the hydrolysis of anandamide more rapidly than that of oleamide^{8,9}.

With intact cells, the inhibitory effect of oleamide was dose-dependent, producing marked elevations in [¹⁴C]anandamide levels (Fig. 1b). When [¹⁴C]anandamide (50 μ M) and oleamide (100 μ M) were co-incubated, there was a 75–100% inhibition of [¹⁴C]anandamide hydrolysis with short incubation times (2–5 min; Fig. 1c), under which conditions the physiological activation of CB₁ receptors by anandamide is likely to occur.

We also tested the ability of oleamide to affect the K_i of anandamide in competition binding experiments (against [³H]HU-243, a high-affinity cannabinoid ligand²) to CB₁ in transfected COS-7 cells¹¹. Anandamide alone yielded a K_i of 350 $\pm$ 7 nM. In the presence of the amidase inhibitor phenylmethylsulphonyl fluoride (PMSF; 200 μ M) the observed K_i decreased by roughly one order of magnitude. A similar enhancement in affinity was seen on addition of oleamide (50 μ M). With PMSF and oleamide together, the affinity did not increase any further.

Naturally occurring oleoylethanolamide (the oleic acid analogue of anandamide),

the hybrid structure between anandamide and oleamide, also fails to activate CB₁ (ref. 12), but has similar effects to oleamide in mice (unpublished results) and inhibits anandamide hydrolysis in rat brain microsomes¹².

Our observations, together with previous results, raise the possibility that some oleamide effects (including induction of sleep) may be mediated by anandamide.

Raphael Mechoulam

Ester Fride

Lumir Hanuš

Tzviel Sheskin

Department of Natural Products,
David Bloom Centre for Pharmacy,
Faculty of Medicine, Hebrew University,
Jerusalem 91120, Israel
e-mail: Mechou@yam.suff.cc.huji.ac.il

Tiziana Bisogno

Vincenzo Di Marzo

Istituto per la Chimica di Molecole
di Interesse Biologico,
Via Tolano 6, 80072 Arco Felice,
Naples, Italy

Michael Bayewitch

Zvi Vogel

Department of Neurobiology,
Weizmann Institute of Science,
Rehovot 76100, Israel

Table 1 Behavioural effects of Δ^9 -THC, anandamide and oleamide

Test	ED ₅₀ values			
	Δ^9 -THC	Anandamide	Oleamide	Anandamide + oleamide
Open field (ambulation)	7.0	3.4	5.7	0.8
Ring immobility	2.3	5.0	18.9	2.9
Hypothermia	9.5	3.6	17.5	1.5
Hot plate	10.3	28.0	29.2	3.0

Ten minutes after intraperitoneal injection, mice were tested in four ways to evaluate cannabinoid-induced effects⁶. For each drug, dose-response curves were obtained using 5–7 doses in the range 1–10 mg per kg body mass ($n=4$ –7 mice for each dose). ED₅₀ values were calculated by nonlinear regression (Graphpad Prism). Anandamide + oleamide denotes anandamide in the presence of 7.5 mg oleamide per kg body mass.

- Cravatt, B. F. *et al.* *Science* **268**, 1506–1509 (1995).
- Devane, W. A. *et al.* *Science* **258**, 1946–1949 (1992).
- Fride, E. & Mechoulam, R. *Eur. J. Pharmacol.* **231**, 313–314 (1993).
- Piomelli, D. *Arachidonic Acid in Cell Signaling* 167–195 (Landes Co., Austin, 1996).
- Santucci, V., Storme, J. J., Soubrie, P. & Le Fur, G. *Life Sci.* **58**, 103–110 (1996).
- Martin, B. R. *et al.* *Pharmacol. Biochem. Behav.* **40**, 471–478 (1991).
- Boring, D. L., Berglund, B. A. & Howlett, A. C. *Prostaglandins Leukot. Essent. Fatty Acids* **55**, 207–210 (1996).
- Maurelli, S. *et al.* *FEBS Lett.* **277**, 82 (1995).
- Cravatt, B. F. *et al.* *Nature* **384**, 83–87 (1996).
- Di Marzo, V. *et al.* *Nature* **372**, 686–691 (1994).
- Vogel, Z. *et al.* *J. Neurochem.* **61**, 352–355 (1993).
- di Tomaso, E., Beltramo, M. & Piomelli, D. *Nature* **382**, 677–678 (1996).

How hosts control worms

Nematodes are a major cause of disease and death in humans, domestic animals and wildlife. Understanding why some individuals suffer severely whereas others exposed to the same infection remain healthy may assist in the development of rational and sustainable strategies to control infection. Here, using a quantitative genetic analysis of the parasitic nematode population that had accumulated naturally in lambs, we find no apparent influence of host genetics on nematode numbers but an extremely strong influence on average worm length and fecundity. Our results indicate that in growing lambs the main manifestation of genetic resistance is the control of worm fecundity.

The lambs were all straight-bred Scottish blackface sheep from a commercial farm in southwest Strathclyde. We took faecal samples from the rectum of each lamb in late May when the lambs were 3–5 weeks old and then at four-week intervals. After taking each sample, animals were treated with a broad-spectrum anthelmintic (albendazole sulphoxide; Rycoben, Youngs Animal Health, Leyland, UK), given at the recommended dose rate of 5 mg per kg body mass. We assessed the worm load after death at 6.5 months; the number of lambs studied was 501 between 1992 and 1995. We used standard parasitological methods to estimate egg counts and worm burdens¹. Larval culture and post-mortem analyses indicated that more than 80% of the parasites present were *Ostertagia circumcincta*. At least 25 randomly chosen female *O. circumcincta* were fostered for each sheep by image analysis (Foster-Findlay Associates Ltd).

The distributions of the numbers of *O. circumcincta* adults and larvae were positively skewed and we transformed the data by taking the logarithm of each trait plus one. We estimated host heritabilities and genetic correlations by residual maximum likelihood fitting an animal model with all known pedigree relationships². We calculated standard errors from the second derivative of the log likelihood profile around the parameter estimate.

The host heritability estimates were 0 for the number of fourth-stage larvae, 0.08 ± 0.09 for the number of fifth-stage larvae (sexually immature adults) and 0.14 ± 0.10 for the number of adult worms. There is no evidence in these results that genetic variation in the natural host influences the worm burden. In contrast, the heritability of worm length was remarkably high at 0.62 ± 0.20 . This high heritability indicates that genetic variation in the host accounted for almost twice as much of the variation in mean

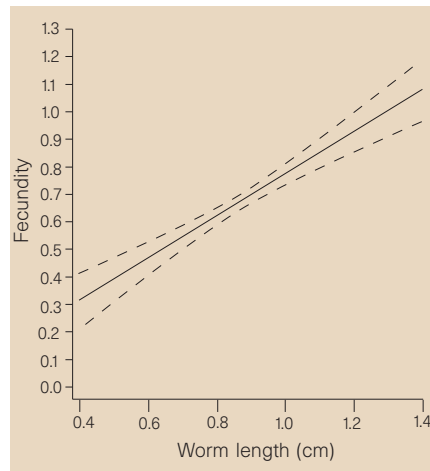


Figure 1 Relationship between worm fecundity and mean worm length. Fecundity was estimated from the log-transformed faecal egg count at slaughter divided by the log-transformed adult worm burden. The 95% confidence limits are shown.

worm length as all other factors combined. There was a strong phenotypic correlation between the number of eggs in the worm uterus and worm length ($r=0.7$; $P<0.0001$) and the heritability of the number of eggs *in utero* was also very high at 0.55 ± 0.19 . An increase in mean worm length is associated with an increase in worm fecundity (Fig. 1). The heritability of faecal egg count at 6 months of age was 0.33 ± 0.14 .

Our results indicate that in a flock of grazing lambs, the genetically resistant hosts³ reduce nematode egg production by decreasing worm fecundity. Although variation in egg output will be influenced by variation in worm burden and worm fecundity, there is no indication that variation in worm burden is under genetic control, at least in lambs.

We find little genetic variation in the number of fourth-stage larvae and suggest that arresting larval development is not an important mechanism of resistance to natural infection in lambs. Previous trials^{4–6} with deliberate infections have indicated that the main immunological mechanism that regulates worm length and fecundity of *O. circumcincta* is the quantity and specificity of parasite-specific immunoglobulin A (IgA). The quantity and specificity of IgA may therefore be an important means of identifying resistant and susceptible lambs.

**M. J. Stear, K. Bairden, J. L. Duncan
P. H. Holmes, Q. A. McKellar, M. Park
S. Strain, M. Murray**

Glasgow University Veterinary School,
Bearsden Road, Bearsden, Glasgow G61 1QH, UK
S. C. Bishop

Division of Biometrical Genetics, Roslin Institute,
Roslin, Midlothian EH25 9PS, UK

G. Gettinby

Department of Statistics and Modelling Science,
University of Strathclyde, Livingston Tower,
26 Richmond Street, Glasgow G1 1XH, UK
e-mail: gyoma05@udcf.gla.ac.uk

1. Armour, J., Jarrett, W. F. H. & Jennings, F. W. *Am. J. Vet. Res.* **27**, 1267–1278 (1966).
2. Meyer, K. *Genet. Sel. Evol.* **21**, 317–330 (1989).
3. Bishop, S. C., Bairden, K., McKellar, Q. A., Park, M. & Stear, M. J. *Anim. Sci.* **63**, 423–428 (1996).
4. Stear, M. J. *et al. Parasite Immunol.* **17**, 643–652 (1995).
5. McCririe, L. *et al. Parasite Immunol.* **19**, 235–242 (1997).
6. Stear, M. J., Park, M. & Bishop, S. C. *Parasitol. Today* **12**, 438–441 (1996).

Salamander with a ballistic tongue

Lungless salamanders of the family Plethodontidae capture prey using the most enhanced tongue-protrusion mechanisms found in amphibians¹. Salamanders of the genus *Hydromantes* are the most extreme specialists^{1–3}, possessing tongues that can be shot with great accuracy at prey several centimetres away (Fig. 1), reaching the target in a few milliseconds. We have found that the tongue of *Hydromantes* is a true projectile. It is fired from the mouth by a ballistic mechanism, and is not simply an extrapolation of the general tongue-protrusion mechanism¹. The tongue reaches the prey under its own momentum.

Species of *Hydromantes* are found in California, Italy and France. They are the only plethodontids, the largest family of salamanders, found in the Old World, all other members being restricted to the Americas^{2,4–6}. As well as the extremely long tongue, *Hydromantes* has other remarkable features including large webbed feet for climbing vertical cave walls and highly sensitive binocular vision which permits detection of prey in as low a light level as 10^{-5} cd m⁻², the illumination level of an overcast, moonless night³.

Like all terrestrial salamanders, *Hydromantes* captures prey on a sticky tongue pad at the anterior tip of a folding tongue, supported by a skeleton of flexible, interlinked cartilages¹ (Figs 1, 2). This skeleton is part of the hyobranchial apparatus, a derivative of the visceral skeleton or gill arches of ancestral vertebrates. The tongue skeleton is driven forwards by protractor muscles acting on its posterior ends, and folds to the midline as it slides forwards (Fig. 2).

In most groups of salamanders, the tongue skeleton is restricted to the floor of the mouth when at rest, but that of *Hydromantes* (and other members of the tribe Bolitoglossini) has tapered and elongated posterior elements that extend over the shoulders and are sheathed for their entire length in circumferential fibres of the tongue protractor muscles¹ (Fig. 2a, b). The protractor muscles are visible as bulges along the sides of the body (Figs 1, 2a) and are unusual in that they exert not only pulling forces as in most salamanders, but also squeezing forces on the elongate elements of the tongue skeleton that launch the tongue from the mouth



Figure 1 Live *H. supramontis* firing its tongue at a housefly. The tongue pad almost engulfs the fly yet is not fully extended. Arrow indicates the posterior extent of the tongue protractor muscle, which is visible as a bulge beneath the skin. Arrowhead points to the posterior ends of the tongue skeleton, which extends forwards to the tongue pad. It has left the mouth entirely and is anchored by the retractor muscles.

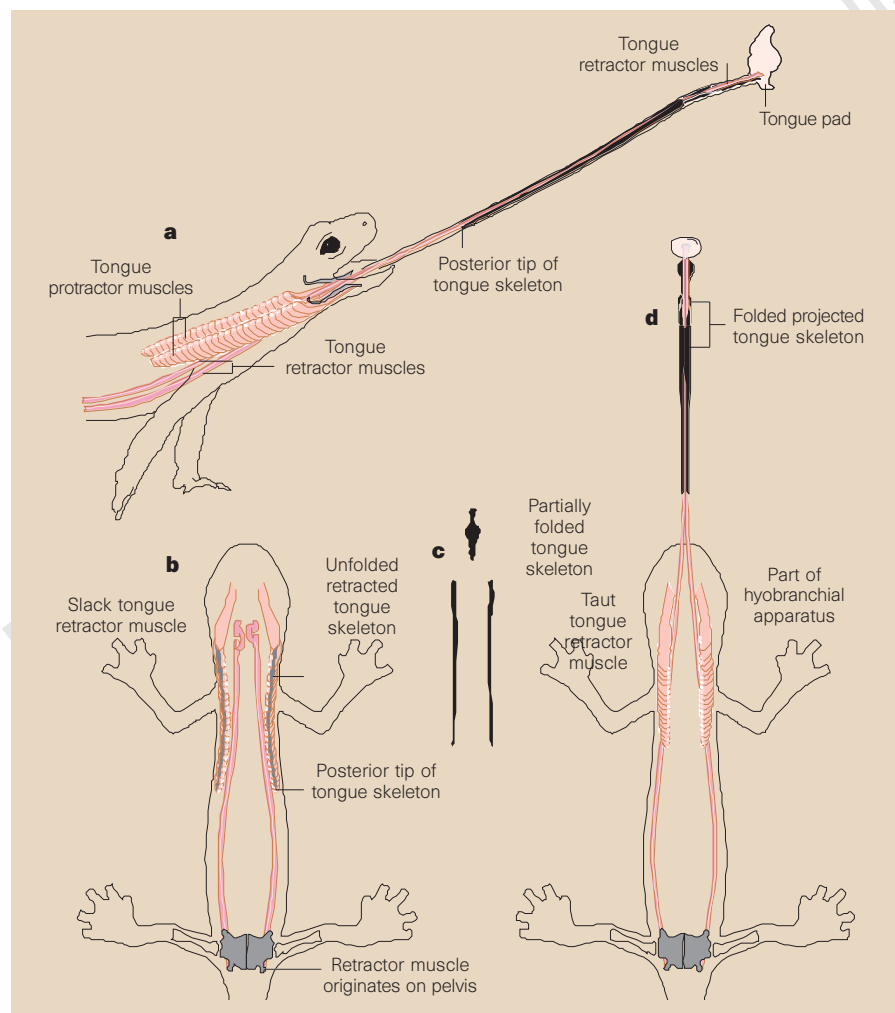


Figure 2 Tongue projection system showing tongue skeleton, associated skeletal elements and the main tongue protractor and retractor muscles. **a**, Lateral view corresponding to Fig. 1 showing the tongue skeleton (black) projected completely from the mouth, and bilaterally paired tongue protractor (peach) and retractor (pink) muscles. **b**, Dorsal view with tongue skeleton in the unfolded retracted position, elongate posterior elements (dark grey) sheathed in protractor muscles, and slack, looped retractor muscles originating on the pelvis (medium grey). **c**, Tongue skeleton partially folded as in an early stage of projection. **d**, Dorsal view showing the tongue skeleton projected entirely from the body, beyond the contracted protractor muscles which are anchored anteriorly to the portion of the hyobranchial apparatus (light grey) that remains in the mouth. The retractor muscles, now taut, run the full length of the extended tongue and body.

(much as a melon seed can be shot from between the fingers).

After the prey is struck, the tongue is returned rapidly to the mouth by lengthy retractor muscles, which in plethodontids originate on the posterior edge of the pelvis and travel uninterrupted into the tongue pad¹ (Fig. 2b, d). The retractor muscles are especially long in *Hydromantes* and other bolitoglossine plethodontids, to accommodate the long-distance protraction. The extra length of muscle is stored as a loop in the throat, just in front of the heart, when the tongue is at rest (Fig. 2b).

We measured preserved museum specimens of *Hydromantes* and found that the length of the folded tongue skeleton is less than 55% of body length. Using a hand-triggered 35 mm camera with strobe, we obtained photographic evidence (Fig. 1) that the tongue of the Sardinian *Hydromantes supramontis* is protruded up to 80% of body length and that the tongue skeleton leaves the mouth completely. The tongue skeleton is projected well beyond the range within which the protractor muscles can impart force on it (Fig. 2a, d), so tongue protraction is truly ballistic.

Ballistic tongue protraction is known in chameleons⁷ and some frogs⁸, but was previously unknown in salamanders. In frogs with ballistic protraction, the tongue has no skeleton⁸ and only soft tissues are projected from the mouth. Only chameleons possess longer projectile tongues than *Hydromantes*, however in chameleons the mechanism is reversed: the hyobranchial apparatus remains in the mouth and the tongue protractor muscle is propelled at the prey⁷. *Hydromantes* launches its entire tongue skeleton at prey and thus is the only vertebrate known to shoot part of its visceral skeleton completely out of its body as a projectile.

Stephen M. Deban

David B. Wake

Museum of Vertebrate Zoology
and Department of Integrative Biology,
3101 Valley Life Sciences Building,
University of California, Berkeley,
California 94720-3160, USA
email: deban@socrates.berkeley.edu

Gerhard Roth

Brain Research Institute,
University of Bremen,
FB 2, D-28334 Bremen, Germany

1. Lombard, R. E. & Wake, D. B. *J. Morphol.* **153**, 39–80 (1977).
2. Dunn, E. R. *The Salamanders of the Family Plethodontidae* (Smith College, Northampton, 1926).
3. Roth, G. *Visual Behavior in Salamanders* (Springer, Berlin, 1987).
4. Lanza, B., Caputo, V., Nascetti, G. & Bullini, L. *Morphologic and Genetic Studies of the European Plethodontid Salamanders: Taxonomic Inferences (Genus Hydromantes)* (Museo Regionale di Scienze Naturali, Torino, 1995).
5. Frost, D. R. *Amphibian Species of the World* (Allen, Lawrence, 1985).
6. Wake, D. B. *Mem. S. Cal. Acad. Sci.* **4**, 1–111 (1966).
7. Wainwright, P. C. & Bennett, A. F. *J. Exp. Biol.* **168**, 23–40 (1992).
8. Nishikawa, K. C. & Gans, C. *J. Exp. Biol.* **199**, 2511–2529 (1996).

Everyday tales of ordinary madness

Why People Believe Weird Things: Pseudoscience, Superstition, and Other Confusions of our Time

by Michael Shermer

W. H. Freeman: 1997. Pp. 306. \$22.95, £16.95

John C. Marshall

Michael Shermer is the California-based publisher of *Skeptic* magazine and director of the Skeptics Society. He is, according to the foreword by Stephen Jay Gould, “an important figure in American life”, defending science against modern irrationalism and pouring justified scorn on those who take advantage of the gullible.

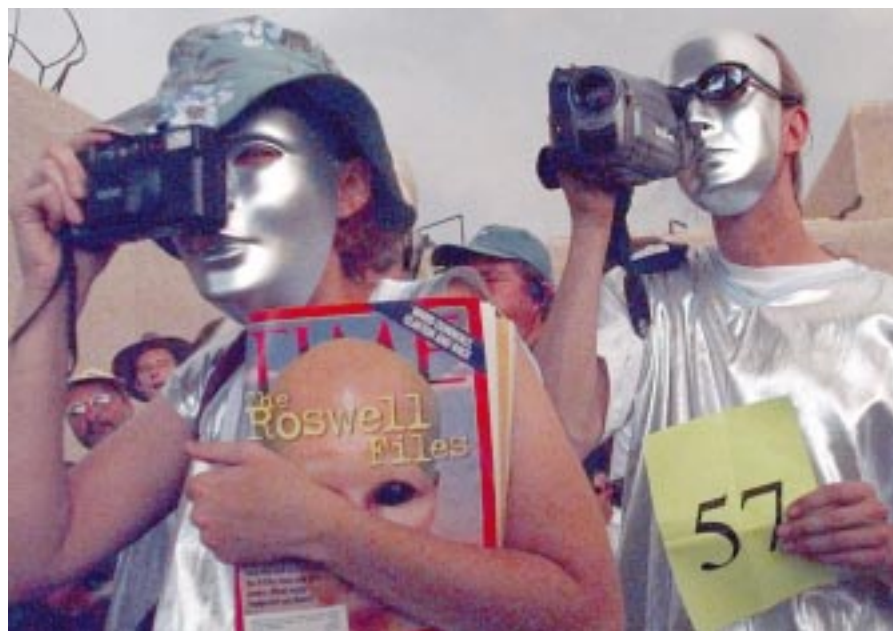
But before we say (a resolutely secular) amen to Shermer’s book, it may be pertinent to cast a sceptical eye on the sceptics. There are many different reasons why people believe weird things and indeed many different weird things that they believe. Shermer rounds up the usual suspects: the paranormal, near-death experiences, alien abduction, Satanism and the ‘recovered memory’ movement, Ayn Rand, and creationism. And what a mixed bag they are.

Shermer seems to have spent a very bad day at Edgar Cayce’s Association for Research and Enlightenment. Cayce (a deceased Kentucky psychic) has already been worked over by Martin Gardner and James Randi. All Shermer adds is that Cayce’s followers seem quite remarkably incompetent at designing and statistically analysing experiments on extrasensory perception (ESP). Would it not have been more revealing to look at some of the many eminently respectable scientists who are conducting kosher research on guessing those little symbols on Zener cards?

On near-death experiences, Shermer cannot, of course, deny that people report leaving their bodies, passing through a tunnel towards the light, and there seeing their loved ones. But the strangely dichotomous question that Shermer then poses — is it more likely that these experiences are “an as-yet-to-be-explained phenomenon of the brain” or evidence of immortality? — is irrelevant to any genuinely scientific issue. Could not both be true?

An acquaintance of mine, an atheist neuroscientist with no belief in (or desire for) immortality, found his near-death experience spiritually rewarding. Perhaps the real question should be: How can arcane brain states be transcendently fulfilling?

Certainly the number of currently unexplained brain states is more than enough to keep neuroscientists busy for some time yet. For example, when friends tell me that they have been abducted by aliens, taken off to UFOs and there experimented upon, I am



Unconventional behaviour: alien costume contest at a meeting held earlier this year to mark the fiftieth anniversary of the “Roswell incident”, the alleged crash-landing of a UFO and its unearthly passengers in New Mexico.

happy to explain that their experiences arise from subclinical temporal lobe epilepsy. But when they ask *how* such overactivity produces those specific experiences, my ums and ers cause them to shake their heads sadly at my credulity.

On satanic abuse, in the Middle Ages it took the rack and the thumbscrew to make people confess that their father drank babies’ blood. Now, a couple of sessions in an air-conditioned psychotherapist’s office seems to suffice. What needs explaining here is how we have become such wimps.

Above all, and despite the title of his book, Shermer refuses to engage deeply with why beliefs are held (or indeed what beliefs, as opposed to scientific hypotheses, actually are). Creationists, I would divine, are lamenting the loss of a monolithic religious culture that never existed. If so, no amount of lecturing from Michael Shermer and Richard Dawkins could reconcile them to their fate, and the latter’s attempts to present science as if it were a religion will only make things worse, on both sides of the divide.

Shermer, at least, appreciates well enough the conjectural nature of science “aimed at building a testable body of knowledge constantly open to rejection or confirmation”. Yet he seems puzzled by the existence of such common beliefs as Holocaust denial and racism. Surely it is only too depressingly obvious why some people need to believe that the Holocaust never happened, or that people with darker skins than their own are intrinsically stupid?

Similarly, he describes objectivism as “the unlikeliest cult in history” when one would have thought it tailor-made for its time and place. Alisa Rosenbaum (aka Ayn Rand) emigrated from post-revolutionary Lenin-grad to New York. There she wrote a best-seller, *The Fountainhead*, that preached the ethics of self-interest and the politics of unbridled capitalism. She gathered around herself a coterie of handsome young men who somewhat assuaged her vociferous appetites, and were also called on to spread her fame as “the greatest human being who ever lived”. I cannot imagine what Shermer finds “weird” about this everyday story of lust, money and power.

By contrast, in the very university in which I work, there are people who claim that an occult force linking the Earth and the Moon provokes the tides to ebb and flow, and that minute creatures too small for the naked eye to see are capable of causing disease. Only this week, I heard that pink worms are dining on solid methane at the bottom of the Gulf of Mexico. And that if you send one member of a pair of simultaneously created photons to Bellevue and the other to Bernex (two villages near Geneva), each one knows what the other is doing and itself does likewise.

To believe in ESP and little green men seems fairly tame in comparison. □

John C. Marshall is at the Neuropsychology Unit, University Department of Clinical Neurology, Radcliffe Infirmary, Woodstock Road, Oxford OX2 6HE, UK.

Scales of progress

Air-Breathing Fishes: Evolution, Diversity and Adaptation

by Jeffrey B. Graham
Academic: 1997. Pp. 299. \$79.95, £55

Molecular Systematics of Fishes

edited by Thomas D. Kocher and Carol A. Stepien
Academic: 1997. Pp. 314. \$79.95, £57

John Long

Perhaps the greatest step in vertebrate evolution was the transition from fishes dwelling in an aqueous habitat to tetrapods crawling on land. Many scientists tend to think of the complex skeletal and physiological changes that happened during the geologically short time span involved, yet few are aware of the many inherent changes in the evolution of fishes that set the stage for their invasion of a new habitat.

Study of the anatomy and physiological mechanisms of air-breathing in living fishes gives an insight into the environmental factors that may have driven the first fishes into adapting an air-breathing behaviour. In *Air-Breathing Fishes*, Jeffrey B. Graham outlines the complete biology of how and why some fishes breathe air, investigates possible reasons for how such adaptations may have evolved, and revisits the fish-tetrapod transition from a fresh viewpoint.

Today some 49 families of fishes have representative air-breathers, falling broadly into two behavioural categories, the amphibious air-breathers and the aquatic air-breathers. Although lungfishes are commonly known as typical air-breathing fishes, there is also a great diversity of actinopterygian fishes

which can partially respire subaerially. The first two chapters of the book comprehensively cover the environmental factors affecting air-breathing, the terminology involved and the diversity of living air-breathing fishes. The remaining chapters deal with the anatomy and physiology of air-breathing fishes, specifically the anatomy of respiratory organs, circulatory adaptations, aerial and aquatic gas exchange and metabolic mechanisms, cardiorespiratory control, blood respiratory properties and metabolic adaptations. The book is well illustrated with clear diagrams, good photographs of dissected specimens, tissue sections and some scanning electron micrographs. It is written in a clear style, is well referenced and has a good index. It should have wide appeal for all interested in the anatomy of fishes and their physiology.

Molecular Systematics of Fishes is a collection of 17 papers covering the range of new and improved methods for taxonomic investigation using such molecular techniques as polymerase-chain reaction amplification and DNA sequencing. Such methods are now widely used for comparing populations of living fishes with their neighbours, or for more distant phylogenetic relationships between species in widely distant taxonomic groups. Despite the all-encompassing title, the book only covers the largest living group of fishes, the teleostean fishes. It holds a wealth of valuable information visually well presented by many cladograms and tables. A must for teleost taxonomists and general fans of phylogenetic systematics. □

John Long is in the Department of Vertebrate Palaeontology, Western Australian Museum, Francis Street, Perth, Western Australia 6000.



From the cover of *Fishes of Chesapeake Bay* by E. O. Murdy, R. S. Birdsong and J. A. Musick. Smithsonian Institution Press, \$49.95, £38.95.

New in paperback

Brainstorms: Philosophical Essays on Mind and Psychology

by Daniel C. Dennett
Penguin, £12.50

Where Does the Weirdness Go? Why Quantum Mechanics Is Strange, But Not As Strange As You Think

by David Lindley
Verso, £7.99

Infinity and the Mind: The Science and Philosophy of the Infinite

by Rudy Rucker
Penguin, £8.99

The Second Creation: Makers of the Revolution in Twentieth Century Physics

by Robert P. Crease and Charles C. Mann
Quartet, £12

Children Talk About the Mind

by Karen Bartsch and Henry M. Wellman
OUP, £14.95

Grassland: The History, Biology, Politics, and Promise of the American Prairie

by Richard Manning
Penguin, \$12.95

Humanity's Descent: The Consequences of Ecological Instability

by Rick Potts
Avon, \$14

The Molecular Vision of Life: Caltech, The Rockefeller Foundation, and the Rise of the New Biology

by Lily E. Kay
OUP, £16.95

The Evolving Coast

by Richard A. Davis Jr
W. H. Freeman/Scientific American Library, \$19.95

Size, Function, and History

by William A. Calder III
Dover, £14.95

Time's Arrows Today: Recent Physical and Philosophical Work on the Direction of Time

edited by Steven F. Savitt
CUP, £16.95, \$24.95

A Natural History of Amphibians

by Robert A. Stebbins and Nathan W. Cohen
Princeton University Press, \$19.95, £15

A sense of place

The Mapping of North America

by Philip D. Burden
 Raleigh Publications, 46 Talbot Road,
 Rickmansworth, Hertfordshire WD3 1HE,
 UK (US office: PO Box 16910, Stamford,
 Connecticut 06905); 1996. Pp. 568. \$195,
 £120

Jared M. Diamond

Today, no prudent motorist, sailor, pilot or hiker sets out into unfamiliar terrain without a printed map. We of the twentieth century take this dependence of travel on maps so completely for granted that we forget how recent it is. The first printed maps date only from the 1470s, a mere two decades after Gutenberg's perfection of printing with movable type around 1455. Techniques of mapmaking evolved rapidly thereafter.

So the most revolutionary change in the history of cartography coincides with the most revolutionary change in Europeans' knowledge of world geography, following Christopher Columbus's discovery of the Americas in 1492.

The earliest sketch map of any part of the Americas dates from 1492 or 1493; the earliest preserved printed map of the Americas from 1506. The succession of printed New World maps that followed is triply interesting: it illustrates the development of cartography, the acquisition of European geographic knowledge of an entire hemisphere, and a series of time-machine-like snapshots of Native American and European colonial settlements.

Philip Burden's lavish book assembles and discusses all 410 known printed maps of North America from that earliest one of 1506 up to 1670. For each map, Burden provides a detailed photograph, a description of its publication, an account of the sources and explorations on which it is based, a list of all surviving copies, and an analysis of variants. This last analysis is critical for identifying which particular map was the model and which were copies or derivatives, what are the differences among them, and the relative contributions of new geographic knowledge and copying changes or errors to those differences.

Many of the maps are ones that Burden himself rediscovered in collections; even more of the maps are preserved only as single copies; and many in private collections remain inaccessible to the public except for the reproductions in this book.

Important discoveries are still to be expected, as tantalizingly indicated by an appendix listing "lost maps" to whose former existence Burden found clues.

Through these 410 maps we see vividly the development of geographic knowledge of North America: for instance, the gradual abandonment of long-persisting erroneous



Abraham Ortelius's classic map of the American continents published in Antwerp in 1570.

beliefs that California is an island, that north-west America has a land connection to Siberia, that a strait separates Central America from South America, and that the Amazon River flows northwards rather than eastwards. Initially less obvious, but even more important, is the book's relevance to the fields of anthropology, biology, epidemiology and linguistics. One example of that relevance must suffice.

Burden reproduces and discusses the first printed detailed regional map of the southeastern United States: the map prepared by the Spanish cosmographer-royal Gerónimo de Chaves, published in 1584 in the *Additamentum III* of Abraham Ortelius's *Theatrum Orbis Terrarum*. As the primary printed contemporary map that drew heavily on findings by the expedition of the Spanish explorer Hernando de Soto through the southeast in 1539–43, this document possesses unique significance.

Along with the Fertile Crescent and China, the southeastern United States was one of barely half-a-dozen areas of the world where agriculture arose independently. Between about 2500 BC and AD 1539, that rise of southeastern agriculture fuelled in turn the independent emergence of towns, chiefdoms, complexly structured societies, and the incipient use of metal. So those developments in the southeast are important to any anthropologist and sociologist interested in understanding the evolution of human societies around the world.

But de Soto and his companions proved to be the last as well as the first literate observers to encounter those southeastern societies intact, because the next European penetration of the region did not take place until the late 1600s.

By that time, the southeastern chiefdoms

no longer existed, having been destroyed by European-born epidemic diseases spread from contacts with de Soto and from European visitors to the coast — diseases to which Europeans had acquired genetic and immune resistance through a long history of exposure, but to which Native Americans had no resistance.

C. Hudson, C. DePratter and M. Smith (pp 77–98 in *First Encounters*, eds. J. T. Milanich and S. Milbrath, University of Florida Press, Gainesville, 1989) have reconstructed de Soto's route through the southeast in detail, by correlating the Chaves/Ortelius map and the Spanish archival map underlying it with archaeological evidence of de Soto's passage. That evidence includes datable sixteenth-century European glass beads and skeletons of Indians whose bones show wounds inflicted by European swords and lances. The Chaves/Ortelius map thereby comes to life, as the sole preserved blueprint of an independently evolved human universe depicted at its peak, and on the eve of its disintegration.

The interpreted map shows us the geographic extent of the chiefdoms; relations between chiefdom size, political complexity and social organization that are central in theories of the evolution of human institutions; the fascinatingly wide uninhabited buffer zones (up to 200 km) separating chiefdoms locked in chronic warfare; the pre-collapse distributions of Native American language groups; the westward attenuation of Native American agricultural systems towards the arid zone of Texas; and the eastward attenuation of bison hunting.

All these unique snapshots are important to linguists, anthropologists and biogeographers, while the depopulation of those soci-

eties continues to provoke a lively debate among epidemiologists and students of infectious diseases.

So much can be gleaned from this map, but each of Burden's 409 other maps has its own corresponding context. Although the resulting book is expensive, it gives value for money with its large format (14 x 10 inches), length (568 pages), and lavish illustrations and referencing. Much of the information is unobtainable elsewhere.

Any large institution of knowledge, and individual scholars concerned with North American colonial history and Native American societies, will require the book. For non-specialists it is fun to browse, beautiful to look at, and an inexhaustible source of interest.

The Mapping of North America would be a nice upmarket present to one of your treasured friends — or to yourself. □

Jared M. Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90095-1751, USA.

Winged wonders

The Atlas of Southern African Birds

edited by J. A. Harrison, D. G. Allan, L. G. Underhill, M. Herremans, A. J. Tree, V. Parker and C. J. Brown
BirdLife South Africa, PO Box 34046, Rhodes Gift, 7707, South Africa: 1997. Two volumes. Pp. 1,500. R648

Michael Cherry

This atlas is the result of the largest biodiversity project yet completed on the African continent. Covering 932 bird species, it represents the work of 5,000 field volunteers over a five-year period, with another five years for data-processing, writing and editing the contributions of 60 authors.

The result has been well worth the wait. For 775 species (all but those that are either vagrants or marginal for the subcontinent), there is a distribution map reflecting the reporting rate for every quarter-degree square of the surface area of six countries (South Africa, Botswana, Namibia, Zimbabwe, Lesotho and Swaziland) as well as a breakdown of the reporting rate for each vegetation type.

A particularly useful feature is a monthly graph of distribution and breeding for each of eight zones into which the subcontinent has been divided (albeit arbitrarily). This allows the reader to infer seasonal patterns, which in Africa are highly variable because of differing climatic conditions. For 51 species with marked seasonality in occurrence, an expanded format with six additional seasonal distribution maps is provided. The advantage of such analysis for migrant species is immediately obvious.

More importantly, the breeding season for a given species often differs between trop-



From a plate in *The Birds of Africa: Volume V* edited by E. K. Urban, C. H. Fry and S. Keith with illustrations by Martin Woodcock and Ian Willis. In a review in *Nature* of volume three, C. M. Perrins said that the project "continues on its way to becoming the standard work on African ornithology for the twenty-first century". Academic, \$145, £85.

ical and subtropical climates, or summer and winter rainfall areas, and these analyses provide revealing insight into these patterns. For biologists engaged in, or planning, studies on southern African birds, this work will prove indispensable.

There are some minor faults: the texts for each species, for example, are not similarly enlightening. Of necessity, reporting for all areas has not been uniformly regular, as the density of competent observers is not evenly distributed throughout the subcontinent. But rather than being disingenuous, the editors are at pains to point out, in a quantifiable way, exactly what the limitations of the data are, which provides a valuable insight into what remains to be done.

What is extraordinary about this achievement is that it could never have been accomplished even 25 years ago — there simply would not have been enough volunteers to complete the task of recording the data. The past quarter century has seen a surge of interest in birding in southern Africa, which has allowed this project to come to fruition.

There are two wishes on my list. The first is that someone has the drive to repeat the exercise before too long, so that the effects of development on the region's avifauna, as a flagship group, can be accurately monitored. The second is that it will inspire the production of atlases for other taxa, so that the extraordinary biodiversity of this corner of the Earth, currently experiencing an exponential increase in development, can be better documented. □

Michael Cherry is in the Department of Zoology, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa.

Classic writings

Galileo's Commandment: An Anthology of Great Science Writing

edited by Edmund Blair Bolles
W. H. Freeman, \$26.95

A bumper collection that ranges widely across time and disciplines, from Herodotus on the Nile Valley to Isaac Asimov on death in the laboratory. The editor, who introduces each piece, divides the book into three parts: an opening section in which scientists try to understand science, a long middle section that shows how the scientific imagination attempts to understand nature, and a third section in which writers transform scientific efforts into literary achievements.

Great Essays in Science

edited by Martin Gardner
OUP, £8.99 (pbk)

First published in 1957 and later reissued under the title *The Sacred Beetle and Other Great Essays in Science*, this anthology aims, in Gardner's words, "to spread before the reader, whether his or her interest in science be passionate or mild, a sumptuous feast of great writing — absorbing, thought-disturbing pieces that have something important to say about science and say it forcibly and well". "There are many pleasurable surprises in this highly individual selection", wrote *Nature's* reviewer of an earlier edition.

Essential Classics in Science

CD-ROM edited by John Gribbin and Pat Coyne
Electric Book Company, 20 Cambridge Drive, London SE12 8AJ, UK (tel./fax: +44 (0)181 488 3872; e-mail: pat-coyne@geo2.poptel.org.uk), £19.95

Complete and unabridged works by Darwin (*Origin of Species*), Einstein (*The Meaning of Relativity*), Faraday (*The Chemical History of a Candle*), Lyell (*Elements of Geology*), Malthus (*Essay on the Principle of Population*), Maxwell (*Matter and Motion*) and Wallace (*The Malay Archipelago*).

Darwin, second edition

CD-ROM
Lightbinders, 2325 3rd St, Suite 324, San Francisco, California 94107, USA (tel: +1 415 621 5746; fax: +1 415 621 5898; e-mail: darwin@lbin.com), \$49.95

Reproduces all of Darwin's major works, a selection of his lesser-known short papers and his 1,200-page monograph on barnacles. Also included are an extensive timeline, biographical dictionary and bibliography of 1,500 primary and secondary sources as well as the text of Michael T. Ghiselin's *Triumph of the Darwinian Method*.

New Journals

Next week's issue will contain *Nature's* annual New Journals review supplement.

The origin and early evolution of plants on land

Paul Kenrick & Peter R. Crane

The origin and early evolution of land plants in the mid-Palaeozoic era, between about 480 and 360 million years ago, was an important event in the history of life, with far-reaching consequences for the evolution of terrestrial organisms and global environments. A recent surge of interest, catalysed by palaeobotanical discoveries and advances in the systematics of living plants, provides a revised perspective on the evolution of early land plants and suggests new directions for future research.

The origin and early diversification of land plants marks an interval of unparalleled innovation in the history of plant life. From a simple plant body consisting of only a few cells, land plants (liverworts, hornworts, mosses and vascular plants) evolved an elaborate two-phase life cycle and an extraordinary array of complex organs and tissue systems. Specialized sexual organs (gametangia), stems with an intricate fluid transport mechanism (vascular tissue), structural tissues (such as wood), epidermal structures for respiratory gas exchange (stomates), leaves and roots of various kinds, diverse spore-bearing organs (sporangia), seeds and the tree habit had all evolved by the end of the Devonian period. These and other innovations led to the initial assembly of plant-dominated terrestrial ecosystems, and had a great effect on the global environment.

Early ideas on the origin of land plants were based on living groups, but since the discovery of exceptionally well-preserved fossil plants in the Early Devonian Rhynie Chert, research has focused almost exclusively on the fossil record of vascular plants^{1,2}. During the 1970s, syntheses of palaeobotanical and stratigraphic data emphasized the Late Silurian and Devonian periods as the critical interval during which the initial diversification of vascular plants occurred^{1,2}, and identified a group of simple fossils (rhyniophytes, such as *Cooksonia* and *Rhynia*) as the likely ancestral forms². They also supported earlier hypotheses of two main lines of evolution: one comprising clubmosses (Fig. 1f) and extinct relatives, the other including all other living vascular plants (ferns, horsetails and seed plants; Fig. 1g–j) and related fossils^{1,2}. During the past two decades, the discovery of fossil spores from as far back as the mid-Ordovician period³, improved knowledge of living green algae^{4,5}, renewed interest in the phylogenetic position of other relevant groups such as mosses and liverworts⁵, and advances in molecular systematics^{6–14}, together with unexpected new data on the structure and biology of Silurian and Devonian fossils^{15–25}, have provided a broader perspective on the origin of a land flora²⁶. These new data indicate that the early diversification of land plants substantially pre-dates the Late Silurian to Early Devonian, and suggest that the main basal lineages originated over a period of more than 100 million years (Myr).

Patterns in the early fossil record

Evidence on the origin and diversification of land plants has come mainly from dispersed spores and megafossils. Gray recognized three new plant-based epochs (Eoembryophytic, Eotracheophytic and Eutracheophytic) spanning the origin and early establishment of land plants: each is characterized by the relative abundance of spore types and megafossils³. This synthesis highlights diversification and floral change in the Ordovician and Silurian^{3,27,28}, and emphasizes a major discrepancy between evidence from spores and megafossils: unequivocal land plant megafossils are first recognized in the fossil record roughly 50 Myr after the appearance of land plant spores.

Eoembryophytic (mid-Ordovician [early Llanvirn: ~476 Myr] to Early Silurian [late Llandovery: ~432 Myr])³. Spore tetrads (comprising four membrane-bound spores; Fig. 2d) appear over a broad geographic area in the mid-Ordovician and provide the first good evidence of land plants^{3,26,29}. The combination of a decay-resistant wall (implying the presence of sporopollenin) and tetrahedral configuration (implying haploid meiotic products) is diagnostic of land plants. The precise relationships of the spore producers within land plants are controversial, but evidence of tetrads and other spore types (such as dyads) in Late Silurian and Devonian megafossils^{16,30}, as well as data on spore wall ultrastructure²⁵ and the structure of fossil cuticles³¹, support previous suggestions of a land flora of liverwort-like plants (Fig. 1c)³. Some early spores and cuticles may also represent extinct transitional lineages between charophycean algae (Fig. 1a, b) and liverworts (Box 1), but precise understanding of their affinities is hindered by the dearth of associated megafossils.

Eotracheophytic (Early Silurian [latest Llandovery: ~432 Myr] to Early Devonian [mid Lochkovian: ~402 Myr])³. The early Silurian (latest Llandovery) marks the beginning of a decline in diversity of tetrads and a rise to dominance of individually dispersed, simple spores, which are found in several basal land plant groups (such as hornworts, some mosses, and early vascular plants)³. Although tetrads remain dominant in some Early Devonian localities from northwestern Europe³², the elaboration of simple spores and turnover of spore 'species'³³ provide evidence of increasing land plant diversity and vegetational change. Although spores have been observed in Silurian megafossils, the affinities of most dispersed forms remain unknown, indicating that substantial land plant diversity is currently undetected in the megafossil record³⁰.

The earliest unequivocal land plant megafossils are from the mid-Silurian of northern Europe³³, and lowermost Upper Silurian of Bolivia³⁴ and Australia³⁵, and the uppermost Silurian of northwestern China³⁶. Early assemblages include clubmosses (such as *Baragwanathia*) and related early fossils (such as zosterophylls, some species of *Cooksonia*), and various other plants of uncertain affinity (such as *Salopella* and *Hedeia*; Fig. 3). These data document an influx into land plant communities of diverse but generally small (usually less than 10 cm tall) organisms related to vascular plants (Fig. 3). Exceptions to the generally small size include the clubmoss *Baragwanathia*³⁷ and the large and much-branched *Pinnatiramosus* from the early Silurian of China³⁸. The habit and branching of *Pinnatiramosus* is similar to that of green algae in the Caulerpales, but the presence of tracheid-like tubes is inconsistent with this interpretation³⁹. Additional details, including conclusive data on reproductive structures, are needed to clarify the relationships of this enigmatic plant.

Data from northern Europe, Siberia, Podolia (southwestern Ukraine), Libya, Vietnam, Bolivia, Australia and Xinjiang and Yunnan (China) document increasing land plant diversity into

the base of the Devonian^{33–36,40}. These fossils, together with the relative chronology implicit in current hypotheses of relationship, imply a minimum mid-Silurian origin for several important vascular plant groups (Box 1; Fig. 4).

Eutracheophytic (Early Devonian [late Lochkovian: ~398 Myr] to mid-Permian [~256 Myr])³. In the Early Devonian (late Lochkovian) the diversity of spores and megafossils increased dramatically^{29,40–42}. Early assemblages include the classic floras from the Rhynie Chert^{20,43,44}, the Gaspé Peninsula of eastern Canada^{43,44}, New York State^{43,44}, the Rhine Valley of Germany⁴⁵, Belgium⁴⁶, Australia³⁵ and Yunnan Province (China)³³, which document a substantial increase in vascular plant diversity, including the appearance and early diversification of many important living groups.

Building a land plant

Phylogenetic studies favour a single origin of land plants from charophycean green algae (Box 1). Based on the ecology of living species, a freshwater origin of land plants seems likely, but direct evidence from the fossil record is inconclusive as mid-Palaeozoic charophytes are found in both freshwater and, more commonly, marine facies⁴⁷. Living charophycean algae (Fig. 1a, b) possess several biosynthetic attributes that are expressed more fully

among land plants, including the capacity to produce sporopollenin, cutin, phenolic compounds and the glycolate oxidase pathway^{4,48}. However, the absence of well-developed sporophytes, gametophytes with sexual organs of land plant type, cuticle and non-motile, airborne, sporopollenin-walled spores suggests that these innovations evolved during the transition to the land^{4,18}. In contrast to animal groups, the entire multicellular diploid phase of the plant life cycle probably evolved in a terrestrial setting.

The transition from an aqueous to a gaseous medium exposed plants to new physical conditions that resulted in key physiological and structural changes. Important metabolic pathways leading to lignins, flavonoids, cutins and plant hormones in vascular plants probably arose from pre-existing elements of primary metabolism in charophycean algae and bryophytes⁴. Although the evolution of these pathways is poorly understood, possible phenolic precursors have been detected in charophycean algae^{4,31}, and elements of auxin metabolism have been recognized in mosses and hornworts⁴⁹.

Phylogenetic studies predict that early land plants were small and morphologically simple, and this hypothesis is borne out by fossil evidence (Fig. 3). Early fossils bear a strong resemblance to the simple spore-producing phase of living mosses and liverworts (Fig. 1d, e and 5)^{16, 45, 51}, and these similarities extend to the anatomical details of the spore-bearing organs and the vascular system¹⁹. The

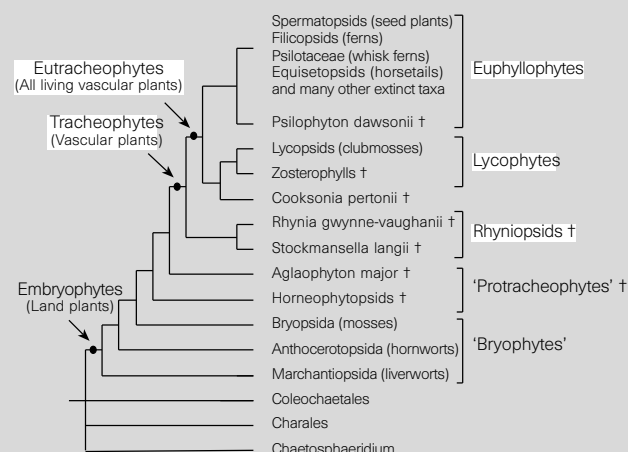
Box 1 Relationships among land plants

Land plants (embryophytes) are most closely related to the Charophyceae, a small group of predominantly freshwater green algae, within which either Coleochaetales (~15 living species; Fig. 1a) or Charales (~400 living species; Fig. 1b), or a group containing both, is sister group to land plants^{4,5,10,12,74}.

Land-plant monophyly is supported by comparative morphology^{4,5,26,75} and gene sequences (18S rRNA, mitochondrial DNA: *cox III*)^{12,14}. Relationships among the major basal living groups are uncertain^{4,5,26,76,77}, but the best-supported hypothesis resolves liverworts (Fig. 1c) as basal and either mosses (Fig. 1e) or hornworts (Fig. 1d) as the living sister group to vascular plants (tracheophytes)^{4,5,13,14,26,75}. Less parsimonious hypotheses recognize bryophyte monophyly and either a sister-group relationship with vascular plants²⁶ or an origin from within basal vascular plants^{14,76,78}.

Among vascular plants, living ferns (Fig. 1g), horsetails (Fig. 1i) and seed plants (Fig. 1j) (euphyllophytes) are the sister group to clubmosses (Fig. 1f)^{13,14,26,75,79}. Euphyllophyte monophyly is strongly supported by comparative morphology²⁶ and a unique 30-kb inversion in the chloroplast genome⁹, as well as sequence data from 18S rRNA¹³ and mitochondrial DNA (*cox III*)¹⁴. These data also provide evidence that the enigmatic Psilotaceae (Fig. 1h) (a group of simple plants once thought to be living relicts of the earliest vascular plants) are more closely related to the fern–seed plant lineage than to basal vascular plants (clubmosses or the extinct rhyniophytes). Within vascular plants, molecular and morphological assessments of phylogeny at the level of orders and below give similar results¹¹, but at deeper levels (for example, the divergence of major groups of ferns, horsetails and seed plants) phylogenetic resolution is poor. These difficulties highlight the problems of approaches based solely on living species⁷⁸⁰. Combined analyses of molecular sequences from multiple loci, and large-scale structural characteristics of the genome (such as introns and inversions), may be more useful in assessing deep phylogenetic patterns.

Megafossils fill some of the substantial morphological ‘gaps’ among living groups. Phylogenetic analyses^{19,26} interpolate two Early Devonian Rhynie Chert plants, *Aglaophyton* and *Horneophyton*, between bryophytes and basal vascular plants as they possess some features unique to vascular plants (a branched, nutritionally independent sporophyte) but also retain bryophyte-like characteristics (terminal sporangia, columella in *Horneophyton*, and the absence of leaves, roots and tracheids with well-defined thickenings). The discovery of previously unrecognized diversity in



extinct *Cooksonia* and similar early fossils (such as *Tortilicaulis*, *Uskiella*, *Caia*^{15–17,81}) (Fig. 3) suggests that simple early land plants (once grouped as rhyniophytes¹) are an unnatural assemblage²⁶. Some *Cooksonia* species may be among the precursors to vascular plants (protracheophytes), whereas others are vascular plants apparently allied to the clubmoss lineage²⁶.

Clubmosses emerge from a poorly resolved grade of extinct *Zosterophyllum*-like plants (Fig. 4), although most zosterophylls form a monophyletic group²⁶. Within clubmosses, early leafy herbaceous fossils such as *Baragwanathia* and *Asteroxylon* are basal^{26,82}, and living Lycopodiaceae are resolved as sister group to a clade that comprises the extinct herbaceous Protolpidodendrales, living *Selaginella* and the predominantly arborescent carboniferous lepidodendrids, including living *Isoetes*^{26,82} (Fig. 4).

Euphyllophytes make up more than 99% of living vascular plants and exhibit much greater diversity than lycophytes. Relationships among basal euphyllophytes are still poorly understood²⁶. Further progress requires a better understanding of the relationships of several fossil groups of uncertain status (such as Trimerophytina, Cladoxylaes and Zygopteridales)^{26,79}.

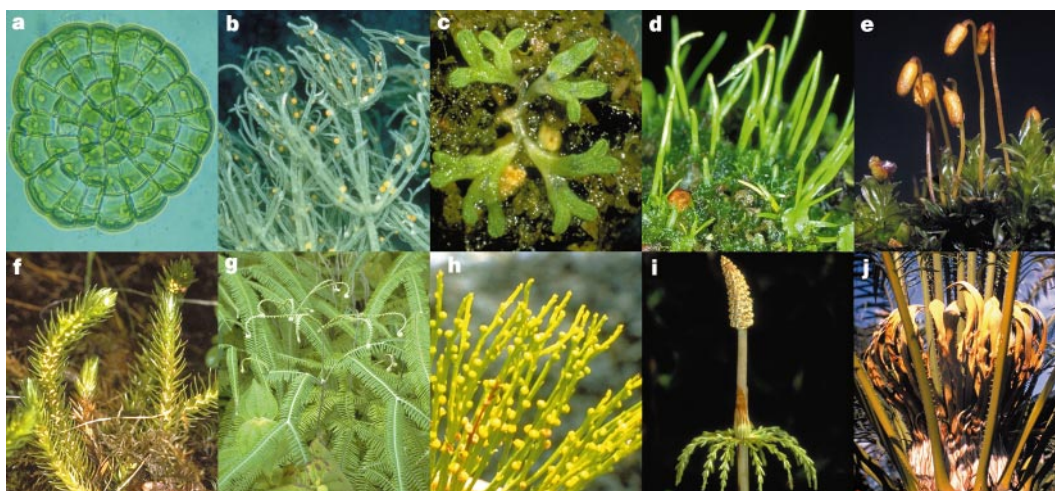


Figure 1 Morphological diversity among basal living land plants and potential land-plant sister groups. **a**, *Coleochaete orbicularis* (Charophyceae) gametophyte; magnification $\times 75$ (photograph courtesy of L. E. Graham). **b**, *Chara* (Charophyceae) gametophyte; magnification $\times 1.5$ (photograph courtesy of M. Feist). **c**, *Riccia* (liverwort) gametophyte showing sporangia (black) embedded in the thallus; magnification $\times 5$ (photograph courtesy of A. N. Drinnan). **d**, *Anthoceros* (hornwort) gametophyte showing unbranched sporophytes; magnification $\times 2.5$ (photograph courtesy of A. N. Drinnan). **e**, *Mnium* (moss) gametophyte showing unbranched sporophytes with terminal sporangia (cap-

sule); magnification $\times 4.5$ (photograph courtesy of W. Burger). **f**, *Huperzia* (clubmoss) sporophyte with leaves showing sessile yellow sporangia; magnification $\times 0.8$. **g**, *Dicranopteris* (fern) sporophyte showing leaves with circinate vernation; magnification $\times 0.08$. **h**, *Psilotum* (whisk fern) sporophyte with reduced leaves and spherical synangia (three fused sporangia); magnification $\times 0.4$. **i**, *Equisetum* (horsetail) sporophyte with whorled branches, reduced leaves, and a terminal cone; magnification $\times 0.4$. **j**, *Cycas* (seed plant) sporophyte showing leaves and terminal cone with seeds; magnification $\times 0.05$ (photograph courtesy of W. Burger).

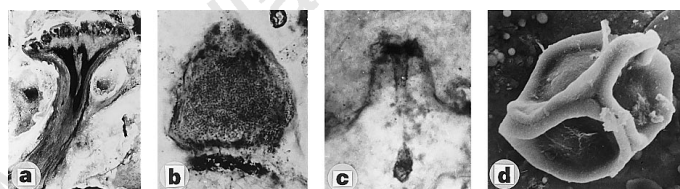


Figure 2 **a**, Longitudinal section of part of a silicified early fossil gametophyte (*Kidstonophyton discoides* from the Rhynie Chert). Antheridia (male sexual organs) are located on the upper surface of the branch; magnification $\times 3.4$. **b**, Longitudinal section of antheridium of *Lyonophyton rhyniensis* from the Rhynie Chert; magnification $\times 40$. **c**, Longitudinal section of archegonium (female sexual organ) of *Langiophyton mackiei* from the Rhynie Chert; magnification $\times 80$. **a-c**

are from the Remy Collection (slides 200, 90 and 330), Abteilung Paläobotanik, Westfälische Wilhelms-Universität, Münster, Germany (photographs courtesy of H. Hass and H. Kerp). **d**, Scanning electron micrograph of *Tetrahedraletes medinensis* showing a spore tetrad of possible liverwort affinity from the Late Ordovician (photograph courtesy of W. A. Taylor); magnification $\times 670$.

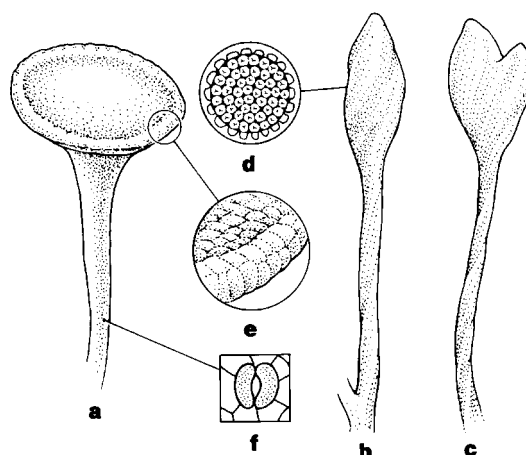


Figure 3 Sporophyte diversity in Early Devonian rhyniophyte fossils. **a**, *Cooksonia pertonii* *apiculispora*: sporophyte (incomplete proximally) with terminal sporangium¹⁵; magnification $\times 15$. **b**, *Tortilicaulis offaeus*: sporophyte (incomplete proximally) with terminal sporangium⁸¹; magnification $\times 40$. **c**, *Tortilicaulis offaeus*: sporophyte (incomplete proximally) with terminal bifurcating sporangium⁸¹; mag-

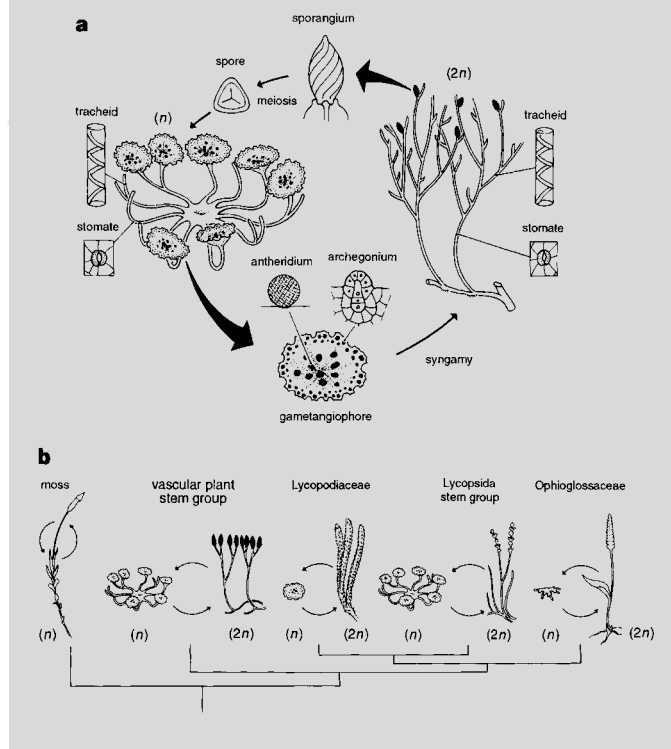
nification $\times 30$. **d**, Transverse section of sporangium showing thick wall and central spore mass; magnification $\times 70$. **e**, Details of epidermis at rim of sporangium; magnification $\times 45$. **f**, Stomate with two reniform guard cells (stippled); magnification $\times 120$.

Box 2 Early evolution of the land plant life cycle

Land-plant life cycles are characterized by alternating multicellular sexual (haploid gametophyte, n) and asexual phases (diploid sporophyte, $2n$). Phylogenetic studies bearing indicate that land plants inherited a multicellular gametophyte from their algal ancestors but that the sporophyte evolved during the transition to the land. Most megafossils are sporophytes, and until recently there was no direct early fossil evidence for the gametophyte phase. Recent discoveries of gametophytes in the Rhynie Chert (Early Devonian, 380–408 Myr) have shed new light on the evolution of land-plant life cycles^{18,20}.

Early gametophytes (**a** in figure) are more complex than in living plants and have branched stems bearing sexual organs on terminal cup- or shield-shaped structures (Fig. 2a). Archegonia (female gametangia) are flask-shaped with a neck canal and egg chamber, and are sunken as in hornworts and most vascular plants (Fig. 2c). Antheridia (male gametangia) are roughly spherical, sessile or with a poorly-defined stalk, and superficial (Fig. 2b). Gametophytes are very similar to associated sporophytes, and shared anatomical features (water-conducting tissues, epidermal patterns, and stomates) have been used to link corresponding elements of the life cycle^{18,20}. Our provisional reconstruction of the life cycle of an early vascular plant is based on information from anatomically preserved plants and contemporaneous compression fossils.

The similarities between gametophyte and sporophyte in early fossil vascular plants contrast strongly with the marked dissimilarities typical of living land plants (**b** in figure). The phylogenetic position of fossils suggests that, after the development of a simple, unbranched, 'parasitic' sporophyte among early land colonizers at the bryophyte grade (such as mosses) there was elaboration of both gametophyte and sporophyte in vascular plants. The implications for interpreting life cycles in living vascular plants^{18,26} are shown. The small, simple, often subterranean and saprophytic gametophytes of living clubmosses (such as Lycopodiaceae) and ferns (such as Psilotaceae, Stromatopteridaceae, Ophioglossaceae) result from morphological loss. Phylogenetic evidence indicates that gametophyte reduction was independent in clubmosses and the fern-seed plant lineage. These data provide a new interpretation of the gametophyte morphology of living clubmosses (Lycopodiaceae)¹⁸.



fossil record also documents significant differences from living groups, particularly in life cycles and the early evolution of the sexual phase (Box 2).

In common with some animal groups, internalization of vital functions and organs (such as gas exchange surfaces and sexual organs), combined with the development of impermeable exterior surfaces, seem to have been primary responses to life on land. Together, these changes resulted in more highly differentiated plants with stomates, multicellular sexual and spore-bearing organs, water-conducting and other tissue systems^{52–54}. Morphological differentiation occurred in both phases of the life cycle (gametophyte and sporophyte), but there was subsequently a dramatic reduction in the gametophyte and a great increase in sporophyte complexity among vascular plants (Box 2). Apical growth and branching coupled with delayed initiation of spore-bearing organs were important innovations of vascular plants that led to a more complex architectural framework on which subsequent morphological diversification was based. The fossil record clearly shows that many vascular-plant organs can be interpreted in terms of modification (especially duplication and sterilization) of basic structural units such as the spore-bearing tissues and the stem^{26,54}. In ferns and seed plants, much morphological diversity is clearly attributable to modifications of branching systems into a variety of leaf-like organs, whereas the relatively conservative clubmoss bauplan has a dearth of organ systems that can be interpreted as modified branches. In both lineages, however, meristem dormancy and abortion were early innovations, providing evidence of hormonal control and substantial phenotypic flexibility^{21,26}.

Early terrestrial ecosystems

The advent of land plants had important consequences for energy and nutrient fluxes among terrestrial and freshwater ecosystems^{29,55} and hence for the evolution of animal groups that live in these habitats. The vegetational changes of the Silurian and Devonian also had a major impact on the atmosphere and other aspects of the global environment. The evolution of roots is thought to have been an important factor in the reduction of atmospheric CO₂ concentrations through increased weathering of Ca–Mg silicate minerals brought about by mechanical disruption and soil acidification^{56,57}. Accelerated weathering has also been linked to the formation of Devonian and Early Carboniferous marine black shales⁵⁸, but this requires further investigation in view of similar deposits earlier and later in the geological record. Root-like impressions have been recognized in Late Silurian palaeosols⁵⁹, but the earliest unequivocal evidence comes from Early Devonian vascular plants²⁶, which have modified prostrate stems bearing rhizoids resembling those of living bryophytes. More substantial roots capable of anchoring large trees evolved independently in several groups during the Middle to Late Devonian.

A further series of innovations in vascular plants, including the biosynthesis of lignin and the origin of lateral meristems (cambium), were critical to the development of large plants, and these developments may have been stimulated by competition for light. Trees evolved independently in several major groups, resulting in stratified forest communities by the end of the Middle Devonian and the production of large amounts of highly decay-resistant organic material (in the form of lignified wood). The early evolution of lignin-decomposing fungi (some Ascomycetes, and Basidiomycetes) is still poorly understood²⁴, but these groups would have been essential for recycling much of the organic carbon.

The earliest land plants probably encountered terrestrial ecosystems that had been occupied by bacteria and protists^{60,61}, algae⁴, lichens^{23,62} and fungi²⁴ since the Late Proterozoic. A variety of enigmatic plants (such as *Protosalvinia*^{44,63}) were also present, and some of the largest elements (*Prototaxites* 'trunks' >69 cm in diameter) may have been fungi⁶⁴. Such organisms, or perhaps some rhyniophytes¹⁶, may be the source of the microscopic tubular

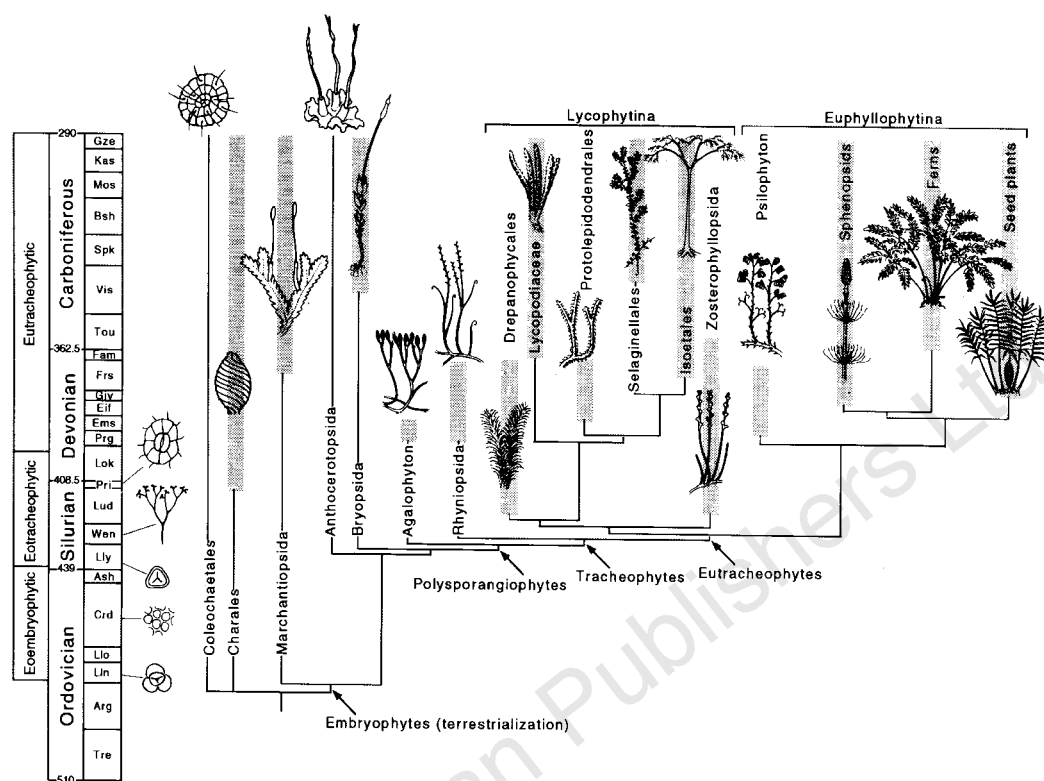


Figure 4 Simplified phylogenetic tree showing the minimum stratigraphic ranges of selected groups based on megafossils (thick bars) and their minimum implied range extensions (thin lines). Also illustrated alongside time scale are minimum age estimates for the appearance of certain important land-plant features (from the bottom: spore tetrads, cuticles, single trilete spores, megafossils and stomates). The first unequivocal record of charophycean algae is based on calcified charalean oogonia (female sexual organs) from the Late Silurian (Pridoli, ~410 Myr)⁸³ and distinctive gametophytes from the Early Devonian Rhynie Chert⁴⁴. Proposed similarities between living *Coleochaete* and Early Devonian *Parka*

remain to be confirmed⁴⁴. Note that megafossil evidence for vascular plants precedes megafossil evidence of bryophytes and charophycean algae. Confirmation that the Early Devonian *Sporogonites* is a plant at the bryophyte grade could help to reduce this discrepancy. Tre, Tremadoc; Arg, Arenig; Lln, Llanvirm; Llo, Llandeilo; Crd, Caradoc; Ash, Ashgill; Lly, Llandovery; Wen, Wenlock; Lud, Ludlow; Pri, Pridoli; Lok, Lochkovian (Gedinnian); Prg, Pragian (Siegenian); Ems, Emsian; Eif, Eifelian; Giv, Givetian; Frs, Frasnian; Fam, Famennian; Tou, Tournaisian; Vis, Visean; Spk, Serpukhovian; Bsh, Bashkirian; Mos, Moscovian; Kas, Kasimovian; Gze, Gzelian.

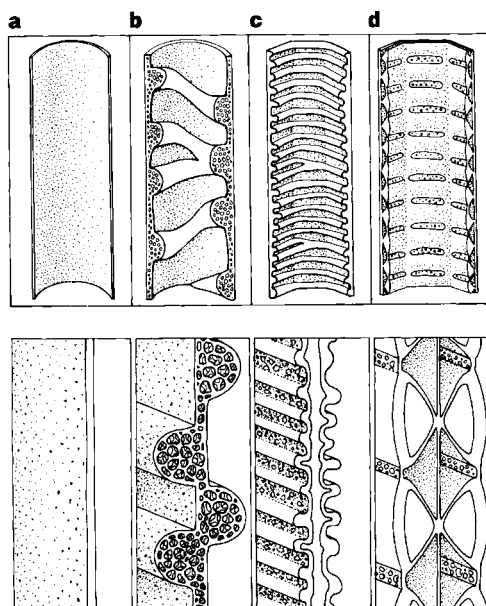


Figure 5 Diversity of water-conducting cells (tracheids) in early land plants (median longitudinal section through cells, basal and proximal end walls not shown; cells are ~20–40 μ m diameter). **a**, Top, bryophyte hydroid; bottom, details of hydroid wall showing distribution of plasmodesmata-derived micropores (10–50 nm diameter; stipple)⁸⁴. **b**, Top, S-type tracheid (fossil) of Rhyniopsida; bottom, details of S-type cell wall showing distribution of plasmodesmata-derived micropores (stipple) and 'spongy' interior to thickenings¹⁹. **c**, Top, G-type tracheid (fossil) of basal extinct eutricheophytes, which closely resemble the tracheids of some living vascular plants; bottom, details of G-type cell wall showing pores distributed between thickenings¹⁹. **d**, Top, scalariform pitted P-type tracheid (fossil) typical of trimerophyte grade plants (euphyllophytes); bottom, details of P-type cell wall showing pit chambers and sheet with pores that extends over pit apertures²⁶.

fragments commonly extracted from Silurian and Early Devonian sediments²⁸. These tubes are often associated with cellular cuticular fragments (*Nematothallus* and *Cosmochlaina*) that may represent fragmented cuticular material from bryophyte-like plants³¹. The discovery of fungal arbusculae in Early Devonian megafossils²² confirms earlier suggestions that endomycorrhizal associations were an important innovation in the colonization of the land⁶⁵.

In contrast to megascopic plants, which appear to have colonized the land only once, many animal groups made the transition to terrestrial existence independently and overcame the problems of water relations in different ways^{52,66,67}. Early evidence for terrestrial animals is sparse^{29,67–69}, but by the Early Devonian exquisitely preserved arthropod faunas are known from several localities in North America, Germany and the United Kingdom^{29,66,67}. These faunas document the appearance of diverse arthropod communities including centipedes, millipedes, trigonotarbids and their living relatives spiders, pseudoscorpions, mites (oribatids and endostigmatids), arthropleurids (extinct arthropods), archaeognathans (primitive wingless insects), collembolans and possibly bristletails. Available evidence indicates that these animals were mainly predators and detritivores and, until the appearance of vertebrate herbivores in the latest Palaeozoic, most energy flow into animal components of early terrestrial ecosystems was probably through the decomposer pathway rather than direct herbivory²⁹. Indirect evidence for herbivory comes from wound responses in the tissues of some fossil plants^{70,71}, and perhaps also from fossil faecal pellets containing abundant spores^{70,72}.

Future directions

The fossil record of spores, combined with phylogenetic studies, indicates that groups related to living bryophytes were early colonisers of the land, and suggests that several major lineages of vascular plant had already evolved by the mid-Silurian. Megafossils of land plants, however, appear much later, and in these assemblages there is a conspicuous bias toward the recognition and perhaps representation of vascular plants. The most important source of data on early megafossils has been the northern European (Laurussian) region, but the appearance of megafossils in this area coincides with facies changes driven by a widespread marine regression^{28,73}, and all Silurian land-plant megafossils are from marine sediments³³. It seems likely that the onset of continental conditions in the Devonian of northern Europe allowed megafossils to be preserved at a time when vascular plants were well established but still diversifying. The rapid appearance of vascular plants in this region^{40–42} owes as much to changing geological conditions as to rapid biological diversification^{27,28}. Intensified sampling in areas that are remote from these regional events is therefore a high priority.

Palaeobotanical evidence shows that the major groups of living land plants are relicts, even though much modern species diversity within these groups may have evolved more recently. Data from the fossil record are therefore especially important for clarifying homologies among major organ systems which may otherwise be difficult to detect as a result of morphological divergence and extinction. Such combined studies of living and fossil plants provide an improved basis for comparative studies of plant development. They indicate, for example, that the ontogeny of leaves and spore-bearing organs in clubmosses are likely to share substantial similarities, but are unlikely to exhibit common features with leaves in seed plants, ferns and horsetails. They also suggest that fundamental features of land plants, such as the spore-bearing organs, stems, stomates and sexual organs, are each under the same kind of developmental control in all main groups. To explore these issues further, data are needed on the molecular basis of plant development from a broader selection of land plants than are currently under study. In the context of a more complete understanding of plant diversity than that provided by living plants alone, such data

should be expected to confirm the underlying unity and relative simplicity of developmental processes in land plants. □

Paul Kenrick is at the Swedish Museum of Natural History, Box 50007, S-104 05, Stockholm, Sweden; Peter R. Crane is at The Field Museum, Roosevelt Road at Lake Shore Drive, Chicago, Illinois 60605, and the Department of the Geophysical Sciences, University of Chicago, USA.

1. Banks, H. P. Reclassification of Psilophyta. *Taxon* **24**, 401–413 (1975).
2. Chaloner, W. G. & Sheerin, A. in *The Devonian System* (eds House, M. R., Scrutton, C. T. & Bassett, M. G.) 145–161 (The Palaeontological Association, London, 1979).
3. Gray, J. Major Paleozoic land plant evolutionary bio-events. *Palaeogeog. Palaeoclimatol. Palaeocol.* **104**, 153–169 (1993).
4. Graham, L. E. *Origin of Land Plants* (Wiley, New York, 1993).
5. Mishler, B. D. et al. Phylogenetic relationships of the "green algae" and "bryophytes". *Ann. MO Bot. Gard.* **81**, 451–483 (1994).
6. Manhart, J. R. & Palmer, J. G. The gain of two chloroplast tRNA introns marks the green algal ancestors of land plants. *Nature* **345**, 268–270 (1990).
7. Manhart, J. R. Phylogenetic analysis of green plant *rbcL* sequences. *Mol. Phylogenet. Evol.* **3**, 114–127 (1994).
8. Raubeson, L. A. & Jansen, R. K. Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. *Science* **255**, 1697–1699 (1992).
9. Chapman, R. L. & Buchheim, M. A. Ribosomal RNA gene sequences: analysis and significance in the phylogeny and taxonomy of green algae. *Crit. Rev. Plant Sci.* **10**, 343–368 (1991).
10. McCourt, R. M., Karol, K. G., Guerlesquin, M. & Feist, M. Phylogeny of extant genera in the family Characeae (Charales, Charophyceae) based on *rbcL* sequences and morphology. *Am. J. Bot.* **83**, 125–131 (1996).
11. Pryer, K. M., Smith, A. R. & Skog, J. E. Phylogenetic relationships of extant ferns based on evidence from morphology and *rbcL* sequences. *Am. Fern J.* **85**, 205–282 (1995).
12. Kranz, H. D. et al. The origin of land plants: phylogenetic relationships among charophytes, bryophytes, and vascular plants inferred from complete small-subunit ribosomal RNA gene sequences. *J. Mol. Evol.* **41**, 74–84 (1995).
13. Kranz, H. D. & Huss, V. A. R. Molecular evolution of pteridophytes and their relationships to seed plants: evidence from complete 18S rRNA gene sequences. *Plant Syst. Evol.* **202**, 1–11 (1996).
14. Hieseler, R., von Haeseler, A. & Brennicke, A. Plant mitochondrial nucleic acid sequences as a tool for phylogenetic analysis. *Proc. Natl Acad. Sci. USA* **91**, 634–638 (1994).
15. Edwards, D., Davies, K. L. & Axe, L. A vascular conducting strand in the early land plant *Cooksonia*. *Nature* **357**, 683–685 (1992).
16. Edwards, D., Duckett, J. G. & Richardson, J. B. Hepatic characters in the earliest land plants. *Nature* **374**, 635–636 (1995).
17. Fanning, U., Edwards, D. & Richardson, J. B. A diverse assemblage of early land plants from the Lower Devonian of the Welsh Borderland. *Bot. J. Linn. Soc.* **109**, 161–188 (1992).
18. Kenrick, P. Alternation of generations in land plants: new phylogenetic and morphological evidence. *Biol. Rev.* **69**, 293–330 (1994).
19. Kenrick, P. & Crane, P. R. Water-conducting cells in early fossil land plants: implications for the early evolution of tracheophytes. *Bot. Gaz.* **152**, 335–356 (1991).
20. Remy, W., Gensel, P. G. & Hass, H. The gametophyte generation of some early Devonian land plants. *Int. J. Plant Sci.* **154**, 35–58 (1993).
21. Remy, W. & Hass, H. New information on gametophytes and sporophytes of *Aglaophyton major* and inferences about possible environmental adaptations. *Rev. Palaeobot. Palynol.* **90**, 175–194 (1996).
22. Remy, W., Taylor, T. N., Hass, H. & Kerp, H. Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proc. Natl Acad. Sci. USA* **91**, 11841–11843 (1994).
23. Stein, W. E., Harmon, G. D. & Hueber, F. M. in *International Workshop on the Biology and Evolutionary Implications on Early Devonian Plants* (Westfälische Wilhelms-Universität Münster, Germany, 1994).
24. Taylor, T. N. & Osborne, J. M. The importance of fungi in shaping the paleoecosystem. *Rev. Palaeobot. Palynol.* **90**, 249–262 (1996).
25. Taylor, W. A. Ultrastructure of lower Paleozoic dyads from southern Ohio. *Rev. Palaeobot. Palynol.* **92**, 269–280 (1996).
26. Kenrick, P. & Crane, P. R. *The Origin and Early Diversification of Land Plants: A Cladistic Study* (Smithsonian Institution Press, Washington DC, 1997).
27. Gray, J. The microfossil record of early land plants: advances in understanding of early terrestrialization, 1970–1984. *Phil. Trans. R. Soc. Lond. B* **309**, 167–195 (1985).
28. Gray, J. & Boucot, A. J. Early vascular land plants: proof and conjecture. *Lethaia* **10**, 145–174 (1977).
29. DiMichele, W. A. et al. in *Terrestrial Ecosystems Through Time: Evolutionary Paleocology of Terrestrial Plants and Animals* (ed. Behrensmeyer, A. K.) 205–325 (Univ. Chicago Press, 1992).
30. Fanning, U., Richardson, J. B. & Edwards, D. in *Pollen and Spores* (eds Blackmore, S. & Barnes, S. H.) 25–47 (Clarendon, Oxford, 1991).
31. Kroken, S. B., Graham, L. E. & Cook, M. E. Occurrence and evolutionary significance of resistant cell walls in charophytes and bryophytes. *Am. J. Bot.* **83**, 1241–1254 (1996).
32. Wellman, C. H. & Richardson, J. B. Sporomorph assemblages from the 'Lower Old Red Sandstone' of Lorne, Scotland. *Special Papers Palaeontol.* **55**, 41–101 (1996).
33. Edwards, D. in *Paleozoic Palaeogeography and Biogeography* (eds McKerrrow, W. S. & Scotese, C. R.) 233–242 (Geological Society, London, 1990).
34. Morel, E., Edwards, D. & Iñiguez Rodríguez, M. The first record of *Cooksonia* from South America in the Silurian rocks of Bolivia. *Geol. Mag.* **132**, 449–452 (1995).
35. Tims, J. D. & Chambers, T. C. Rhyniophytina and Trimerophytina from the early land flora of Victoria, Australia. *Palaeontology* **27**, 265–279 (1984).
36. Cai, C.-Y., Dou, Y.-W. & Edwards, D. New observations on a Pridoli plant assemblage from north Xinjiang, northwest China, with comments on its evolutionary and palaeogeographical significance. *Geol. Mag.* **130**, 155–170 (1993).
37. Hueber, F. M. Thoughts on the early lycopods and zosterophylls. *Ann. MO Bot. Gard.* **79**, 474–499 (1992).
38. Cai, C. et al. An early Silurian vascular plant. *Nature* **379**, 592 (1996).
39. Geng, B.-Y. Anatomy and morphology of *Pinnatinnosus*, a new plant from the Middle Silurian (Wenlockian) of China. *Acta Bot. Sin.* **28**, 664–670 (1986).
40. Raymond, A. & Metz, C. Laurussian land-plant diversity during the Silurian and Devonian: mass extinction, sampling bias, or both? *Paleobiology* **21**, 74–91 (1995).
41. Edwards, D. & Davies, M. S. in *Major evolutionary radiations* (eds Taylor, P. D. & Larwood, G. P.) 351–376 (Clarendon, Oxford, 1990).
42. Knoll, A. H., Niklas, K. J., Gensel, P. G. & Tiffney, B. H. Character diversification and patterns of evolution in early vascular plants. *Paleobiology* **10**, 34–47 (1984).
43. Gensel, P. G. & Andrews, H. N. *Plant Life in the Devonian* (Praeger, New York, 1984).

44. Taylor, T. N. & Taylor, E. L. *The Biology and Evolution of Fossil Plants* (Prentice Hall, New Jersey, 1993).
45. Schweitzer, H.-J. Die Unterdevonflora des Rheinlandes. *Palaeontographica B* **189**, 1–138 (1983).
46. Gerrienne, P. Inventaire des végétaux éodévonien de Belgique. *Ann. Soc. Géol. Belg.* **116**, 105–117 (1993).
47. Tappan, H. N. *The Paleobiology of Plant Protists* (Freeman, San Francisco, 1980).
48. Raven, J. Plant responses to high O₂ concentrations: relevance to previous high O₂ episodes. *Palaeogeog. Palaeoclimatol. Palaeocol.* **97**, 19–38 (1991).
49. Sztein, A. E., Cohen, J. D., Slovin, J. P. & Cooke, T. J. Auxin metabolism in representative land plants. *Am. J. Bot.* **82**, 1514–1521 (1995).
50. Edwards, D. New insights into early land ecosystems: a glimpse of a Lilliputian world. *Rev. Palaeobot. Palynol.* **90**, 159–174 (1996).
51. Edwards, D., Fanning, U. & Richardson, J. B. Stomata and sterome in early land plants. *Nature* **323**, 438–440 (1986).
52. Raven, J. A. Comparative physiology of plant and arthropod land adaptation. *Phil. Trans. R. Soc. Lond. B* **309**, 273–288 (1985).
53. Raven, J. A. The evolution of vascular plants in relation to quantitative functioning of dead water-conducting cells and stomata. *Biol. Rev.* **68**, 337–363 (1993).
54. Niklas, K. J. *Plant Allometry: The Scaling of Form and Process*. (Univ. Chicago Press, 1994).
55. Beerbower, R. in *Geological Factors and the Evolution of Plants* (ed. Tiffney, B. H.) 47–92 (Yale Univ. Press, New Haven, CT, 1985).
56. Berner, R. A. GEOCARB II: a revised model of atmospheric CO₂ over Phanerozoic time. *Am. J. Sci.* **294**, 56–91 (1994).
57. Mora, C. I., Driese, S. G. & Colarusso, L. A. Middle to Late Paleozoic atmospheric CO₂ levels from soil carbonate and organic matter. *Science* **271**, 1105–1107 (1996).
58. Algeo, T. J., Berner, R., Maynard, J. B. & Scheckler, S. E. Late Devonian oceanic anoxic events and biotic crises: “rooted” in the evolution of vascular land plants? *GSA Today* **5**, 45, 64–66 (1995).
59. Retallack, G. J. in *Paleosols: their Recognition and Interpretation* (ed. Wright, V. P.) (Blackwell, Oxford, 1986).
60. Knoll, A. H. The early evolution of eukaryotes: a geological perspective. *Science* **256**, 622–627 (1992).
61. Bengtson, S. (ed) *Early life on Earth*. (Columbia Univ. Press, New York, 1994).
62. Taylor, T. N., Hass, H., Remy, W. & Kerp, H. The oldest fossil lichen. *Nature* **378**, 244 (1995).
63. Hemsley, A. R. in *Ultrastructure of Fossil Spores and Pollen* (eds Kurmann, M. H. & Doyle, J. A.) 1–21 (Royal Botanic Gardens, Kew, 1994).
64. Hueber, F. M. in *International Workshop on the Biology and Evolutionary Implications of Early Devonian Plants* (Westfälische Wilhelms-Universität, Münster, 1994).
65. Simon, L., Bousquet, J., Léveque, C. & Lalonde, M. Origin and diversification of endomycorrhizal fungi with vascular plants. *Nature* **363**, 67–69 (1993).
66. Selden, P. A. & Edwards, D. in *Evolution and the Fossil Record* (eds Allen, K. C. & Briggs, D. E. G.) 122–152 (Belhaven, London, 1989).
67. Gray, J. & Shear, W. Early life on land. *Am. Sci.* **80**, 444–456 (1992).
68. Gray, J. & Boucot, A. J. Early Silurian nonmarine animal remains and the nature of the early continental ecosystem. *Acta Palaeontol. Pol.* **38**, 303–328 (1994).
69. Retallack, G. J. & Feakes, C. R. Trace fossil evidence for Late Ordovician animals on land. *Science* **235**, 61–63 (1987).
70. Scott, A. C., Stephenson, J. & Chaloner, W. G. Interaction and coevolution of plants and arthropods during the Palaeozoic and Mesozoic. *Phil. Trans. R. Soc. Lond. B* **336**, 129–165 (1992).
71. Banks, H. P. & Colthart, B. J. Plant-animal-fungal interactions in early Devonian trimerophytes from Gaspé, Canada. *Am. J. Bot.* **80**, 992–1001 (1993).
72. Edwards, D., Seldon, P. A., Richardson, J. B. & Axe, L. Coprolites as evidence for plant-animal interaction in Siluro-Devonian terrestrial ecosystems. *Nature* **377**, 329–331 (1995).
73. Allen, J. R. L. Marine to fresh water: the sedimentology of the interrupted environmental transition (Ludlow-Siegenian) in the Anglo-Welsh region. *Phil. Trans. R. Soc. Lond. B* **309**, 85–104 (1985).
74. Melkonian, M. & Surek, B. Phylogeny of the Chlorophyta: congruence between ultrastructural and molecular evidence. *Bull. Soc. Zool. Fr.* **120**, 191–208 (1995).
75. Bremer, K., Humphries, C. J., Mishler, B. D. & Churchill, S. P. On cladistic relationships in green plants. *Taxon* **36**, 339–349 (1987).
76. Garbary, D. J., Renzaglia, K. S. & Duckett, J. G. The phylogeny of land plants: a cladistic analysis based on male gametogenesis. *Plant Syst. Evol.* **188**, 237–269 (1993).
77. Capesius, I. A molecular phylogeny of bryophytes based on the nuclear encoded 18S rRNA genes. *J. Plant Physiol.* **146**, 59–63 (1995).
78. Taylor, T. N. The origin of land plants: some answers, more questions. *Taxon* **37**, 805–833 (1988).
79. Rothwell, G. W. in *Pteridology in Perspective* (eds Camus, J. M., Gibby, M. & Johns, R. J.) (Royal Botanic Gardens, Kew) (in the press).
80. Albert, V. A. *et al.* Functional constraints and *rbcL* evidence for land plant phylogeny. *Ann. MO Bot. Gard.* **81**, 534–567 (1994).
81. Edwards, D., Fanning, U. & Richardson, J. B. Lower Devonian coalified sporangia from Shropshire: *Salopella* Edwards & Richardson and *Tortilicaulis* Edwards. *Bot. J. Linn. Soc.* **116**, 89–110 (1994).
82. Bateman, R. M., DiMichele, W. A. & Willard, D. A. Experimental cladistic analysis of anatomically preserved lycopods from the Carboniferous of Euramerica: an essay on paleobotanical phylogenetics. *Ann. MO Bot. Gard.* **79**, 500–559 (1992).
83. Feist, M. & Grambast-Fessard, N. in *Calcareous Algae and Stromatolites* (ed. Riding, R.) 189–203 (Springer, Berlin, 1991).
84. Hebert, C. in *Bryophyte Systematics* (eds Clarke, G. C. S. & Duckett, J. G.) 365–383 (Academic, London, 1979).

Acknowledgements. We thank W. G. Chaloner, D. Edwards, J. A. Raven, P. S. Herendeen, E. M. Friis, S. Bengtson and especially J. Gray for criticisms of earlier drafts of this manuscript; W. Burger, J. Cattel, A. N. Drinnan, M. Feist, L. E. Graham, H. Haas, H. Kerp, W. A. Taylor and P. Lidmark for assistance with illustrations. This work was supported in part by the Swedish Natural Science Research Council (NFR) and the National Science Foundation.

Correspondence should be addressed to P.K. (e-mail: pb-paul@nrm.se).

YOURS TO HAVE AND TO HOLD BUT NOT TO COPY

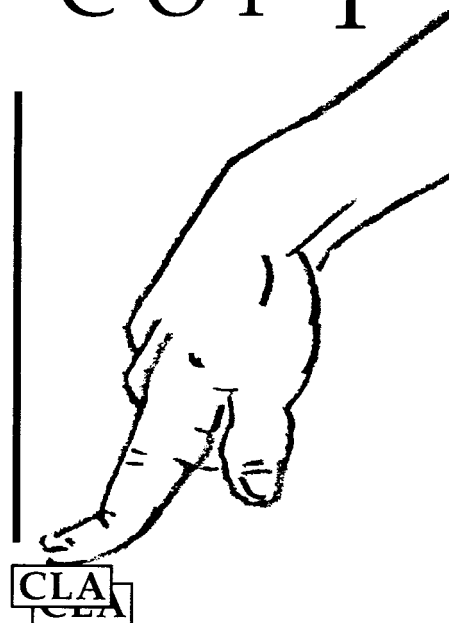
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised infringing photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931
Fax: 0171 436 3986

Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse

Feng Guo, Deshmukh N. Gopaul & Gregory D. Van Duyne

Johnson Research Foundation and Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA

During site-specific DNA recombination, which brings about genetic rearrangement in processes such as viral integration and excision and chromosomal segregation, recombinase enzymes recognize specific DNA sequences and catalyse the reciprocal exchange of DNA strands between these sites. The bacteriophage recombinase Cre catalyses site-specific recombination between two 34-base-pair *loxP* sites. The crystal structure at 2.4 Å resolution of Cre bound to a *loxP* substrate reveals an intermediate in the recombination reaction, in which a Cre molecule has cleaved the substrate to form a covalent 3'-phosphotyrosine linkage with the DNA. Four recombinases and two *loxP* sites form a synapsed structure in which the DNA resembles models of four-way Holliday-junction intermediates. The Cre-*loxP* complex challenges models of site-specific recombination that require large changes in quaternary structure. Subtle allosteric changes at the carboxy termini of the Cre subunits may instead coordinate the cleavage and strand-exchange reactions.

Based on sequence similarity, recombinases from bacteria and yeast have been classified into the integrase family and the resolvase-invertase family, which use two distinct mechanisms of recombination^{1,2}. Integrase family members cleave their DNA substrates by a series of staggered cuts, during which the recombinase becomes covalently linked to the DNA through a catalytic tyrosine residue (Fig. 1). Recombination takes place by exchanging one set of strands to yield a Holliday structure intermediate, followed by the exchange of a second set of strands to resolve the intermediate into recombinant products.

The >60 members of the integrase family³ share poor sequence similarity overall, although four residues required for catalysis are strictly conserved^{4,5}. The sequence dissimilarity among integrase family members may reflect the diversity of function and form in which these recombinases carry out their genetic rearrangements. Some members, such as the bacteriophage-λ integrase, require additional accessory proteins to facilitate the recombination process⁶. In contrast, the Cre recombinase (relative molecular mass, 38 K) from bacteriophage P1 mediates a site-specific recombination reaction between two 34-base-pair *loxP* sites (Fig. 1a, b) which does not require accessory factors and can be performed *in vitro* with a variety of DNA substrates^{7,8}. The role of Cre recombinase in the bacteriophage P1 life cycle is to maintain the phage genome as a monomeric, unit-copy plasmid in the lysogenic state^{7,9}. The simplicity of the Cre/*loxP* system has led to its use in a growing number of *in vitro* and *in vivo* applications^{10,11}, including the engineering of synthetic antibody libraries¹², site-directed chromosome translocations¹³, targeted gene deletions¹⁴, mosaic genomes¹⁵ and gene-specific humanized animal models¹⁶. Advances in this technology have resulted in transgenic systems in which Cre-mediated genetic rearrangements can be controlled in site-specific, tissue-specific and time-specific manners^{17,18}.

The largest gap in our understanding of how Cre recombinase and other integrase family members catalyse the site-specific recombination reaction has been a result of a lack of three-dimensional structural models for the macromolecular complexes involved. However, the three-dimensional structures for the catalytic domains of the bacteriophage λ and HP1 integrases^{3,19} and the structure of the *Escherichia coli* XerD recombinase²⁰ have recently been reported. Using the Cre/*loxP* system as a model, we are

working to bridge this gap by studying the structures of intermediates in the site-specific recombination reaction. We have determined the three-dimensional structure at 2.4 Å resolution of a covalent intermediate in the site-specific recombination reaction between Cre recombinase and a symmetrized, suicide *loxP* substrate. This Cre-DNA complex is a structural model for a site-specific recombination synapse and provides several unexpected mechanistic insights into the recombination reaction.

Structure determination

Cre recombinase was expressed in soluble form in *E. coli* and purified to homogeneity. The purification and construction of DNA substrates for co-crystallization was simplified by replacing the asymmetric 8-base-pair spacer sequence between inverted repeats in *loxP* with the symmetric spacer shown in Fig. 1b. This modified substrate (*loxA*) was constructed from two modular half-sites that dimerize through 4-base-pair 5' overhangs to yield a doubly nicked duplex²¹. The choice of positions for the missing phosphodiester linkages in *loxA* results in the formation of a trapped covalent intermediate following cleavage by Cre. Because the residue on the 3' side of the cleaved phosphate is able to diffuse away, the 5'-hydroxyl group that would normally serve as a nucleophile in the subsequent strand-exchange step of the reaction is eliminated (Fig. 1c). A similar strategy was used to prepare covalent λ integrase-DNA intermediates²². SDS-PAGE analysis of Cre/*loxA* mixtures showed a band of reduced mobility corresponding to formation of a covalent Cre-DNA complex. Crystals of the Cre-*loxA* complex (see Methods), dissolved directly in SDS-containing buffer, showed the same slower-moving band. Crystallographic phases for the Cre-*loxA* structure were determined at 3 Å resolution using a combined multiple isomorphous replacement/anomalous scattering (MIRAS) and multiwavelength anomalous diffraction (MAD) approach using diffraction data measured at the NSLS X25 and X4A beamlines. This combined phasing approach resulted in an experimental electron density map that was readily interpretable, and refinement of the initial model with 2.4 Å data for the parent complex converged rapidly to $R = 0.201$ and $R_{\text{free}} = 0.264$ for 50–2.4 Å resolution data. Data collection statistics, phasing statistics and refinement results are summarized in Table 1.

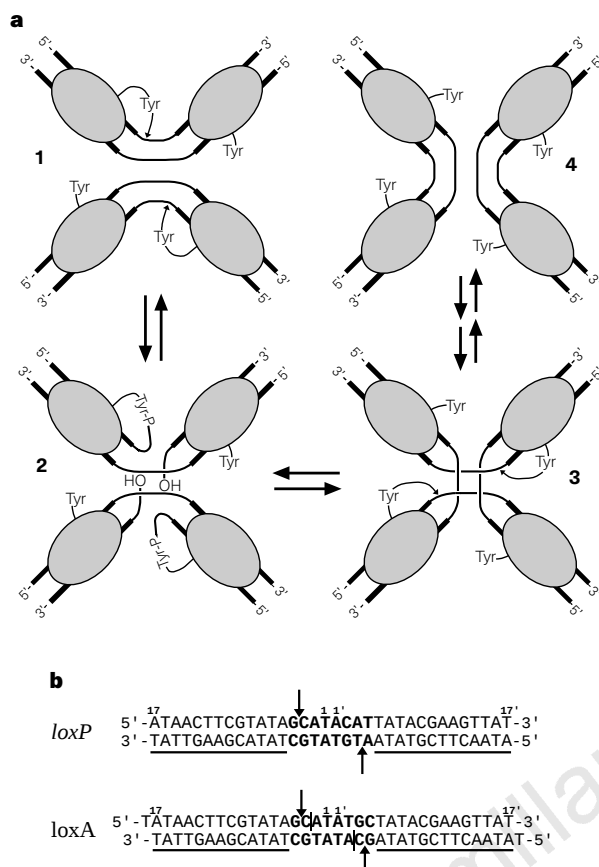


Figure 1 a, The Cre-*loxP* site-specific recombination reaction, based on studies in the lambda integrase family¹ and on the work described here. Two Cre molecules bind to *loxP* sites which then associate to form a recombination synapse (**1**). Conserved Tyr324 cleaves each substrate to form covalent 3'-phosphotyrosine intermediates in an antiparallel arrangement (**2**). (Cleavage of the substrates could precede synapsis.) The 5'-hydroxyl groups released by cleavage of the phosphodiester linkages become nucleophiles in a strand-transfer reaction, in which they attack the phosphotyrosines of the partner substrates, yielding a Holliday-junction intermediate (**3**). A second round of cleavage and strand-exchange reactions gives recombinant products (**4**). The reaction intermediate shown in **2** represents the Cre-*loxA* synapse structure described here. **b**, Sequence of *loxP* and of the DNA substrate (which we have named *loxA*) used to form Cre/DNA cocrystals. The *loxA* construct was formed from 16-mer (5'-TAACTTCGTATAGC-3') and 19-mer (5'-ATATGCTATACGAAGTTAT-3') oligonucleotides that were annealed together. The 13-bp inverted repeats in the *loxP* and *loxA* sites are underlined and the phosphate groups cleaved during the Cre-*lox* recombination reaction are indicated by arrows. The central 8 bp comprise the strand exchange, or crossover region. Vertical lines in the *loxA* sequence identify the location of missing phosphodiester linkages in the top and bottom strands. The number scheme follows refs 26, 30. **c**, The trapped 3'-phosphotyrosine intermediate formed when Cre cleaves the *loxA* substrate. The Cyt3 residue on the 3' side of the cleaved phosphate is able to diffuse away, eliminating the 5'-hydroxyl group that would normally serve as a nucleophile in the subsequent strand-exchange step of the reaction.

Table 1 Summary of crystallographic analysis

Data collection statistics									
Derivative (no. of sites)		Source	Resolution limit (Å)	R_{sym}^*	Redundancy†	Completeness (%)			
Native	$\lambda = 0.9790 \text{ Å}$	NSLS X25‡	2.4	0.069	5.0	93			
Se-Met1§	$\lambda = 0.9840 \text{ Å}$	NSLS X4A	3.0	0.067	2.7	84			
Se-Met2 (22)	$\lambda = 0.9793 \text{ Å}$	NSLS X4A	3.0	0.068	2.7	83			
Se-Met3 (22)	$\lambda = 0.9790 \text{ Å}$	NSLS X4A	3.0	0.068	2.7	83			
Se-Met4 (22)	$\lambda = 0.9790 \text{ Å}$	NSLS X25	2.7	0.041	2.0	97			
Mersalyl (6)		Laboratory	3.0	0.051	2.6	88			
Iodine-1 (4)	$\lambda = 0.9790 \text{ Å}$	NSLS X25	3.0	0.076	4.6	87			
Iodine-2 (4)		Laboratory	3.4	0.073	3.1	88			
Iodine-3 (4)		Laboratory	4.0	0.087	3.7	96			
Phasing statistics									
Resolution limit (Å)	19.8	11.0	7.6	5.8	4.7	4.0	3.4	3.0	All
Phasing power/anomalous phasing power¶									
Se-Met2		1.5/0.93	1.6/0.84	1.6/0.83	1.4/0.72	1.1/0.59	0.83/0.52	0.66/0.44	0.98/0.58
Se-Met3		1.2/0.97	1.1/0.98	1.1/1.0	0.88/0.98	0.66/0.82	0.49/0.72	0.37/0.62	0.59/0.79
Se-Met4	0.43/0.41	0.72/0.77	0.77/0.74	0.93/0.85	0.76/0.85	0.52/0.73	0.45/0.63	0.46/0.54	0.56/0.68
Mersalyl	0.58	1.5	1.6	1.7	1.2	0.83	0.78	0.82	0.96
Iodine-1	0.88	0.76	0.79	0.88	0.62	0.41	0.36	0.37	0.47
Iodine-2	0.61	0.97	1.1	1.1	0.72	0.56	0.51		0.64
Iodine-3	0.66	1.3	1.0	0.96	0.70	0.52			0.70
Mean figure of merit	0.44	0.76	0.77	0.78	0.70	0.58	0.50	0.43	0.56
Refinement results									
	Resolution (Å)		R -factor#		Free R -factor#		No. of reflections		
Data with $F > 2\sigma$	50–2.4		0.201		0.264		41,368		
All data	50–2.4		0.208		0.272		42,896		
R.m.s. deviations									
	Bond lengths 0.005 Å			Bond angles 1.2°		B values (bonded) 3.7 Å			

* $R_{\text{sym}} = \sum |I_h - \langle I_h \rangle| / \sum I_h$, where $\langle I_h \rangle$ is the average intensity over symmetry equivalent measurements.

† Redundancy is the number of measured data/number of unique data.

‡ NSLS X25, NSLS X4A: National Synchrotron Light Source beamlines X25 and X4A, Brookhaven National Laboratory.

§ Se-Met1–3 data were measured from the same crystal in a multiple-wavelength anomalous diffraction experiment. Dispersive differences (treated as isomorphous differences) for Se-Met2 and Se-Met3 are with respect to Se-Met1. Se-Met4 data were measured from a different crystal and isomorphous differences are with respect to the native data.

|| Iodine-1, short strand in Fig. 1b is changed to 5'-TAXAAGCTTCGTATAGC-3', where X is 5-iodo-dU; iodine-2, long strand is changed to 5'-ATATGCTATACGAAGTTAT-3'; iodine-3, long strand is changed to 5'-ATATGCTATACGAAGTTAT-3'.

¶ Phasing power = $\sum |F_H| / \sum |F_{\text{calc}}|$; anomalous phasing power = $\sum |F_H| / \sum |AD_{\text{calc}}|$, where AD is the anomalous difference.

R -factor = $\sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$, where summation is over data used in the refinement; free R -factor includes only the 10% of data excluded from all refinements.

Architecture of the Cre-loxA complex

The structure of the Cre-loxA complex reveals two Cre molecules bound to a single loxA site (Fig. 2). Each Cre molecule contacts the outermost 15 base pairs of one loxA half-site, which includes the 13-base-pair inverted-repeat sequence and the first two base pairs of the central strand-exchange region. One of the two Cre molecules bound to the loxA site has cleaved the DNA substrate to form a covalent 3'-phosphotyrosine linkage. Two antiparallel Cre-loxA complexes are associated in the crystal to form a dyad-symmetric recombination synapse (Fig. 5a). A similar protein-protein interface is formed both between Cre molecules bound to the same loxA site and between Cre molecules bound to different loxA sites in the synapse, generating pseudo-C₄ symmetry in the synaptic assembly. The cleaved loxA sites are bent by ~100° at their centres, resulting in a pseudo-fourfold arrangement of loxA half-site arms in the synapse that resembles a square planar isomer of a Holliday-junction intermediate^{23,24}. The protein:DNA scaffold surrounds a 'strand exchange cavity' which contains the central six bases of the crossover region. Single-stranded DNA segments that are released by the cleavage reaction are not contacted by the recombinase tetramer and have both flexibility and unencumbered access to the active sites of the Cre-loxA synapse.

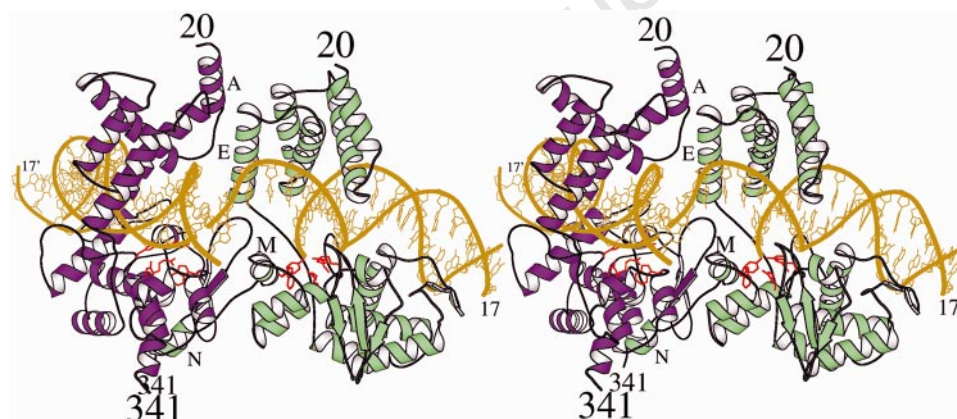


Figure 2 Stereo ribbon model⁴⁹ of the Cre-loxA complex. The Cre subunit that has cleaved the loxA site is in green (right) and the Cre subunit that has not cleaved the loxA site is purple (left). Side chains of the active-site residues (Arg 173, His 289, Arg 292, Trp 315 and Tyr 324) are in red. The N- and C-terminal amino-acid numbers, the first and last bases in the uncleaved DNA strand, and selected α-helices are labelled.

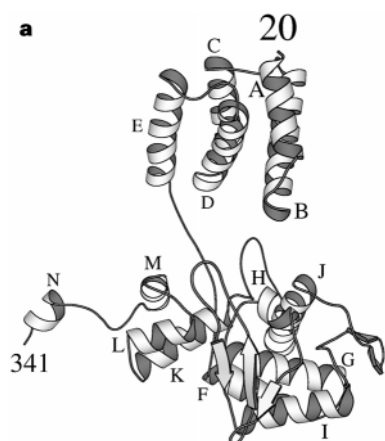


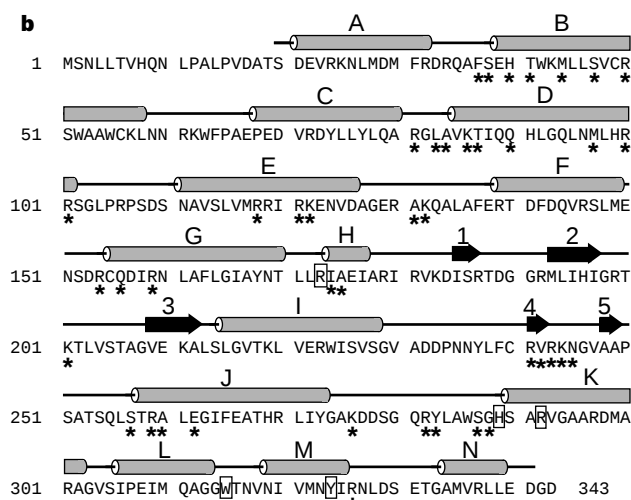
Figure 3 a, Ribbon model⁴⁹ of Cre recombinase. α-Helices are labelled A-N and the amino and carboxy termini are indicated by residue numbers. **b**, Secondary structure assignment for Cre recombinase. Active-site residues are boxed; asterisks indicate amino acids that contact DNA (<3.6 Å) in the subunit that has cleaved the loxA site.

Structure of Cre recombinase

Cre folds into two distinct domains that are separated by a short linker (Fig. 3a, b). The amino-terminal domain (residues 20–129) contains five α-helical segments connected by short loops. Three of the helices (C, D and E) are organized into an antiparallel bundle. Helices A and B are nearly orthogonal to the three-helix bundle, with helix B in close contact with a hydrophobic bundle surface. Helix A is only loosely associated with the rest of the domain. As discussed below, helices A and E are involved in formation of the recombinase tetramer and helices B and D contact the major groove of the loxA half-site. The two independent amino-terminal domain conformations in the Cre-loxA complex are similar, with an r.m.s. deviation of 0.7 Å for α-carbon atoms.

The larger carboxy-terminal domain of Cre recombinase includes amino acids 132–341 and is primarily helical. A small β-sheet (strands 1–3) packs against one surface of a nine-helix domain (helices F–N). The terminal helix N extends away from the rest of the domain and is involved in intersubunit contacts. The C-terminal domain of Cre recombinase is structurally similar to the lambda integrase and HP1 integrase catalytic domains^{3,19}. But unlike the amino-terminal domain, the C-terminal domain differs significantly between the two loxA-bound Cre molecules, with the

Arg 292, Trp 315 and Tyr 324) are in red. The N- and C-terminal amino-acid numbers, the first and last bases in the uncleaved DNA strand, and selected α-helices are labelled.



largest structural deviations coming from the C-terminal segments leading to helix N. The two C-terminal domain structures have r.m.s. deviations of 2.5 Å for all α -carbon atoms and 1.5 Å for α -carbon atoms omitting the final 17 amino acids.

The Cre-*loxA* interface

As shown in Fig. 2, two Cre molecules are bound to the *loxA* site. Each Cre molecule contacts a *loxA* half-site and one of the two molecules has formed a covalent 3'-phosphotyrosine linkage with the DNA. The amino and carboxy-terminal domains of Cre form a clamp around the half-site, making extensive contacts with both the major and minor grooves. This Cre-DNA interface buries $\sim 5,000 \text{ \AA}^2$ of solvent-accessible surface area for each Cre-half-site interaction. The contacts we observe are in good agreement with the results of DNase and hydroxyl radical footprinting studies using both full-length Cre and a Cre deletion mutant lacking the N-terminal domain^{25,26}. The number of electrostatic interactions between Cre and the phosphate backbone of the DNA substrate is striking. A total of 39 contacts to the *loxA* site involving arginine, lysine and histidine side chains are formed by the two Cre subunits. Although the two Cre-DNA interfaces are similar, a clear asymmetry exists that may result in part from conformational changes associated with cleavage of the DNA substrate by only one of the two Cre subunits.

The N-terminal domain contacts the DNA primarily via the orthogonal B and D helices, burying $\sim 1,800 \text{ \AA}^2$ of solvent-accessible surface area in the interface formed by the non-cleaving Cre subunit and $\sim 2,300 \text{ \AA}^2$ in the interface formed by the cleaving subunit. The crossed B and D helices sit in the major groove, centred at positions 5-6 and 5'-6' of the *loxA* sites, and extend about three base pairs in each direction. Together, helix B and helix D make three direct contacts to DNA bases. Helix E of the amino-terminal domain in the

cleaving Cre subunit is aligned directly over the DNA backbone, where it forms three close electrostatic interactions with phosphates at the concave surface of the bent *loxA* site. This interface is different in the non-cleaving subunit, where helix E is instead positioned between adjacent *loxA* substrates in the synapse (Figs 2, 5a).

The C-terminal domain interacts with the entire 13-base-pair inverted-repeat region of the *loxA* half-site plus the first two base pairs of the strand-exchange region. This interface is more complex, involving the entire face of the domain, with both secondary structural elements and connecting loops interacting with the major groove, the minor groove and the sugar-phosphate backbone. The protein surface area buried at this interface is $\sim 3,000 \text{ \AA}^2$ for both the cleaving and non-cleaving subunits, but only a small fraction of this surface involves major-groove interactions. The amino terminus of helix J interacts with the major groove near base pairs 10-11, where Arg 259 forms hydrogen bonds to N7 and O6 of guanine 10 (Gua 10). This is the only direct major-groove contact we observe involving the C-terminal domain. A striking feature of this domain-DNA interface is the number of direct contacts to bases in the minor groove. A short β -strand between helices I and J makes three direct contacts. For example, the side chain of Lys 244 bridges thymine (Thy)-16 O2 and Thy-17 O2. Similarly, the loop between strands 2 and 3 of the β -sheet bridges Gua-4 N3 and Thy-5 O2 with the side chain of Lys 201. A third minor-groove interaction involves Arg 282 in the loop between helices J and K, which forms hydrogen bonds to adenine (Ade)-7 N3.

Active sites and the phospho-Tyr linkage

The Cre active site contains the conserved catalytic triad residues Arg 173, His 289 and Arg 292, the conserved nucleophilic Tyr 324 and Trp 315. The amino acids in a single active site are all derived from the same subunit, indicating that Cre recombinase does not form a shared active site like F1p recombinase²⁷. There are two distinct active sites in the Cre-*loxA* complex, one for each Cre subunit bound to a *loxA* half-site. As the synapse is formed by the association of two Cre-*loxA* complexes, the synapse tetramer contains a total of four active sites, corresponding to the four scissile phosphates in the site-specific recombination reaction. One active site in the Cre-*loxA* complex (two per synapse) contains a 3'-phosphotyrosine DNA-Cre linkage resulting from cleavage of the DNA substrate. This covalently attached phosphate is activated for the strand-exchange step of the reaction, which requires nucleophilic attack of the phosphotyrosine by the 5'-hydroxyl released upon cleavage of the partner substrate (Fig. 1a).

Figure 4a shows the stereochemistry of the active site after cleavage of the DNA substrate. The non-bridging oxygen atoms of the scissile phosphate are coordinated by five hydrogen bonds. Three hydrogen bonds come from the guanidino groups of Arg 173 and Arg 292, which contact the oxygen atoms through their Ne and N η hydrogen atoms. One oxygen atom also receives a hydrogen bond from the Ne hydrogen of His 289, which is ideally positioned to serve as proton donor to the Tyr 324 leaving group during the strand-transfer reaction. An additional hydrogen bond to the second phosphate oxygen atom comes from the side chain of Trp 315. This tryptophan is not conserved in the integrase family, and a histidine residue is more often found in this position. The side chains of histidine and tryptophan could, however, donate similar hydrogen bonds and this functionality may well be conserved in the integrase family. Together, the active-site His, Arg and Trp side chains form a positively charged 'proton cradle' that appears to be preorganized to accommodate the negative charges of equatorial phosphate oxygen atoms in a trigonal bipyramidal transition state of the phosphoryl transfer reaction.

The other active site in the Cre-*loxA* complex surrounds a scissile phosphate that has not been cleaved by Cre (Fig. 4b). A superposition of the two Cre subunits shows that the catalytic arginine side chains and the Trp-315 side chain occupy similar positions in

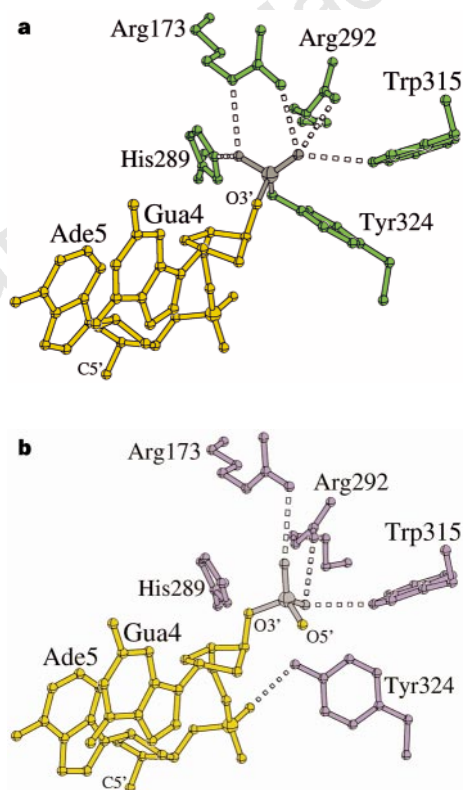


Figure 4 The Cre-*loxA* active sites. Perspective of the active site **a**, where Cre recombinase has cleaved the DNA substrate to form a 3'-phosphotyrosine linkage; and **b**, where Cre has not cleaved the DNA substrate.

this active site. Tyr 324, however, is shifted away from the scissile phosphate with a phosphorus–hydroxyl distance of 5.8 Å. In this conformation, the tyrosine hydroxyl group forms a hydrogen bond with the phosphate group between Gua 4 and Ade 5. This change in structure is partly due to a concerted 3 Å shift in helix M, which contains Tyr 324. His 289 has also shifted away from the scissile phosphate in this active site. We suggest that the stereochemical differences between the two active sites may in part reflect the allosteric consequences of the initial cleavage reaction. The resulting asymmetry is likely to be of mechanistic importance, because the stepwise nature of the integrase family site-specific recombination reaction requires that only one half-site in each substrate be cleaved in preparation for a given exchange of DNA strands.

Interactions in the synaptic complex

The interface between Cre subunits in the synapse tetramer is extensive, with 13,800 Å² of solvent-accessible protein surface buried overall, or 3,450 Å² buried per subunit. The contacts between Cre molecules bound to the same *loxA* site are similar to the contacts between Cre molecules bound to different *loxA* sites, giving rise to the pseudo-fourfold symmetry of the synapse assembly. The N-terminal domain contributes to this network of interactions primarily with helices A and E, which form a parallel helix–helix interface (Figs 2, 5a). The C terminus of helix E also contacts the β-

turn between strands 2 and 3 of the C-terminal domain in the adjacent subunit. The interactions between C-terminal domains in neighbouring subunits are intriguing. The segment following helix M (amino acids 326–333) leads to a short C-terminal helix (N) that buries one face in a hydrophobic pocket of the adjacent subunit (Fig. 2). The hydrophobic pocket is formed partly by the L and M helices, which are integral components of the recombinase active sites. This intermolecular grapple crosslinks Cre subunits in a cyclic manner (Fig. 5c) and may provide a means of communication through which changes in the conformation of helix M and in the segment leading to helix N are relayed to the adjacent subunit. A reciprocal, rather than cyclic, exchange of helices was observed in the structure of the HP1 integrase catalytic domain, where the domains crystallized as dimers that mutually exchanged their C-terminal helix homologues of Cre¹⁹. RuvA, a protein involved in mediating branch migration of Holliday junctions, uses an intermolecular organization in forming a tetramer that is similar to the cyclic exchange observed in the Cre–*loxA* synapse²⁸.

The two Cre subunits bound to the *loxA* substrate differ in the conformation of the segment connecting helix M and helix N. Helix M in the subunit that has cleaved *loxA* is shifted by ~3 Å relative to the same helix in the non-covalently bound subunit, in order to form the phosphotyrosine linkage. As a consequence, the segment connecting helices M and N adopts a more extended conformation

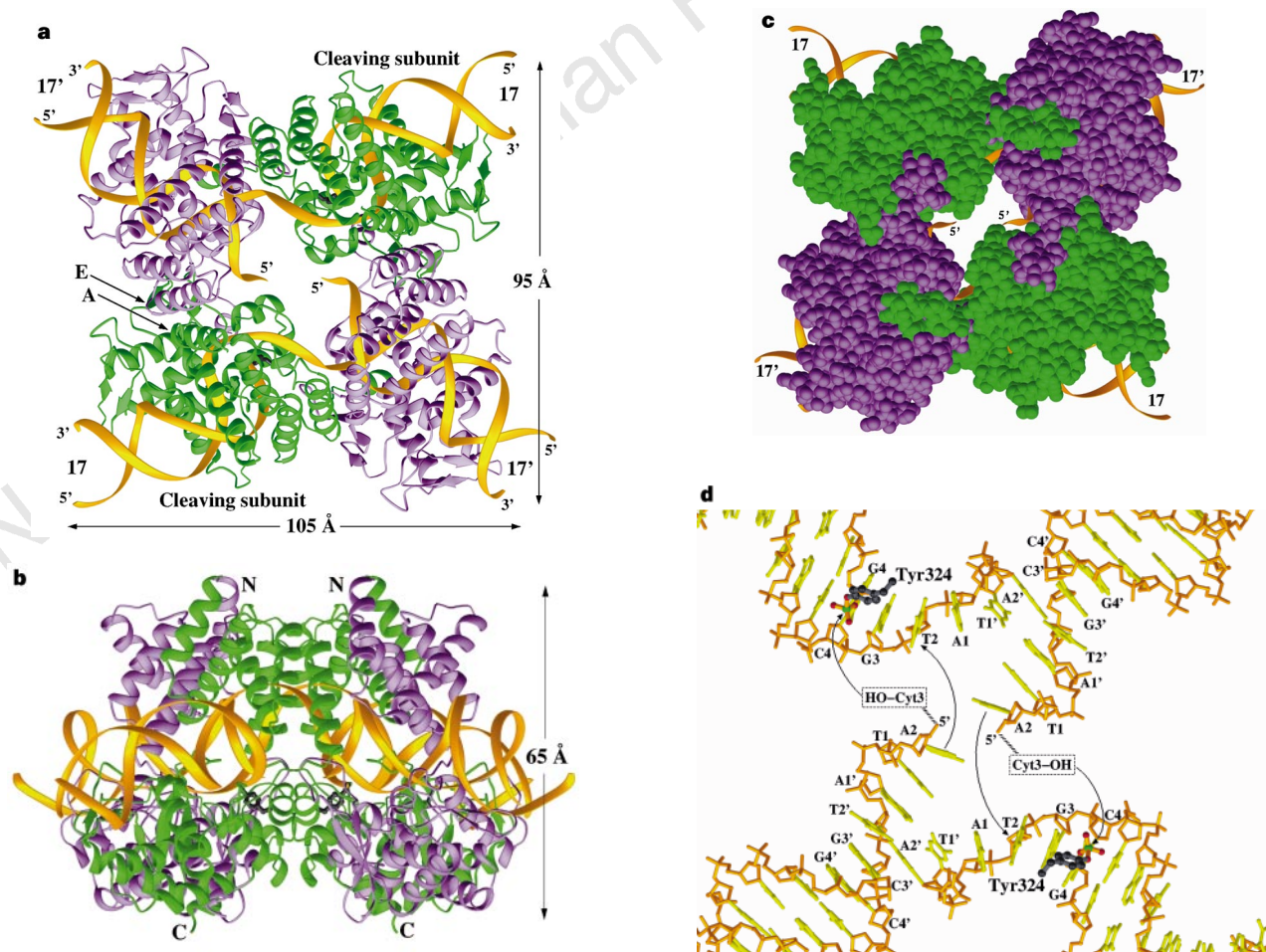


Figure 5 Ribbon and space-filling models⁵⁰ of the Cre–*loxA* synapse. **a**, View from the N-terminal domain side of the tetramer, parallel to the dyad of symmetry and to the pseudo-fourfold axis of symmetry. Helices A and E are indicated. **b**, Side view, rotated 90° about a horizontal axis with respect to **a**. Approximate positions of the N and C termini are indicated. Uncleaved DNA strands are drawn as continuous ribbons in **a** and **b**. **c**, Space-filling model of the recombinase tetramer, showing

the cyclic interaction between subunits involving the C-terminal segments of Cre; the view is from the C-terminal side of the synapse. **d**, Close-up of the DNA substrate in the same orientation as in **c**, with the recombinase subunits omitted. The direction of nucleophilic attack by missing Cyt 3 and formation of a recombinant base-pair between Ade 2 and Thy 2 in the strand-exchange step of the recombination reaction are indicated by arrows. Tyr 324, visible in **b** and **d**, is in black.

in the cleaving subunit (Fig. 2). In the non-cleaving Cre subunit, the position of helix M causes the M–N linker segment to adopt a more compact conformation involving formation of a short 3_{10} helix and a different set of interactions with the cleaving subunit located across the synapse on the partner substrate (Fig. 5c).

Structure of the synapsed *loxA* DNA

The synapsed *loxA* sites have the overall appearance of a distorted four-fold symmetric Holliday structure, with arms formed by the duplex inverted repeat segments of the *loxA* half-sites (Fig. 5a, d). However, as the strand-exchange step of the recombination reaction has been blocked, the covalent bonds leading to a Holliday intermediate have not been formed. There is a smooth bend of $\sim 25^\circ$ within the inverted repeat arms that is present in both the cleaved and uncleaved half-sites. This bend produces a concave DNA surface on one side of the synapse that is evident in Fig. 5b. An increase in the width of the minor groove occurs in the regions involving side chain–minor groove base contacts, but the helical parameters of these duplex segments are otherwise generally consistent with those of B-form DNA²⁹.

The crossover region in the cleaved *loxA* site is composed of both single- and double-stranded segments (Fig. 5d). In the cleaved strand (top strand in Fig. 1b, c), Gua 4 maintains standard base-pairing geometry in spite of the phosphotyrosine linkage. Cytosine 3 has been released by the cleavage reaction and is missing from the electron density, as expected. Ade 2 and Thy 1 are not base-paired with the uncleaved strand, but instead have moved towards the centre of the strand-exchange cavity, where their weak electron density and high thermal parameters indicate a high degree of mobility. The remaining four nucleotides (Ade 1' to Cyt 4') in the cleaved strand are base-paired, with a trend of decreasing mobility and increasing electron density quality nearer to the end of the strand-exchange region. The first two single-stranded bases in the uncleaved strand (Gua 3 and Thy 2) are stacked against neighbouring Cyt 4 and have well-defined conformations. Ade 1 is poorly stacked against its neighbours and provides the nucleation point for a bend in the *loxA* site that leads to a $\sim 100^\circ$ disposition between half-site arms. This base has a poorly defined conformation in the Cre–*loxA* structure and is the only position within the strand-exchange region of *loxP* that was previously found to be intolerant to mutation³⁰.

To investigate whether the framework of the Cre–*loxA* synapse has the appropriate geometry to mediate the strand-exchange reaction, we modelled the missing Cyt 3 to the 5' end of Ade 2 and manually moved the DNA chain towards the phosphotyrosine of the partner substrate (arrows in Fig. 5d). By adjusting their backbone torsion angles, Cyt 3 and Ade 2 could be base-paired with Gua 3 and Thy 2 of the partner substrate, placing the 5'-hydroxyl of Cyt 3 in an appropriate position for nucleophilic attack. Only the torsion angles of Thy 1, Ade 2 and Cyt 3 needed to be changed to adopt this strand-attack geometry. From this analysis, the strand-exchange step of the site-specific recombination reaction could occur in the nucleoprotein architecture of the Cre–*loxA* synapse without major alterations in quaternary structure.

Cre–*loxP* site-specific recombination

Models of the λ integrase family site-specific recombination reaction have until recently been based on the assumption that the branch point of the Holliday junction intermediate is formed at the initial strand-exchange site. To exchange a second set of DNA strands, branch migration and isomerization of the Holliday structure would be required to bring the appropriate pair of cleavage sites into mutual proximity³¹. The requirement for homology between recombining sites has been explained by the need for branch migration of the Holliday junction between cleavage sites, which would stall upon encountering a mismatch^{1,32}. In contrast, we find that the four active sites in the Cre–*loxA* synapse are symmetrically

positioned around their respective scissile phosphates following the first cleavage step of the recombination reaction. In this arrangement, the distances between adjacent active sites are all similar and the *loxA* half-site arms adopt a pseudosquare planar conformation. We therefore propose that the Holliday junction intermediate adopts a nearly four-fold symmetric structure that is stabilized by Cre–DNA interactions and that the cleavage and strand-exchange steps of the recombination reaction occur within the Cre/DNA framework, with no requirement for large-scale protein–DNA isomerization steps (Fig. 1a). A more recent model for site-specific recombination in the λ integrase family requires only limited branch migration of the Holliday junction intermediate, in which the branch point is formed near the centre of the crossover region^{33–35}. Homology between recombining sites is tested in this model by formation of recombinant base pairs during the strand-exchange steps of the reaction. The Cre–*loxA* structure strongly supports this model; the arrows in Fig. 5d indicate how two such recombinant base pairs might be formed.

The stepwise nature of the integrase-family recombination mechanism requires that only one half-site of the DNA substrate be cleaved before strand exchange. What prevents double-strand cleavage? In the synaptic complex, a number of subtle conformational changes involving the C-terminal 40 amino acids may accompany the first strand-cleavage step and prevent attack of the neighbouring half-site through allosteric effects. In the unsynapsed Cre–DNA complex, where two Cre molecules bind to a *loxP* site, cleavage at only one half-site may be ensured by the formation of only a single competent active site. As shown in Fig. 2, only one of the C-terminal segments in a Cre–*loxP* complex would be able to interact with its adjacent Cre subunit, even if the bend in the uncleaved *loxP* site were somewhat different. Without the appropriate intersubunit contacts, the second C-terminal segment may adopt a different structure, or it may not adopt a unique conformation, as seen in the structure of the isolated λ integrase catalytic domain³. It is also unclear how the structure of Cre recombinase would change when bound to an isolated half-site substrate, where there would be no intersubunit contacts. This could be important for interpreting experiments that have relied on the structural integrity of the recombinases when bound to modified DNA half-site probes³⁶.

The three short helices (L–N) that comprise the C-terminal 40 amino acids of Cre also lie at the heart of a mechanistic paradox in the integrase family³⁶. In question is the source of the conserved tyrosine that is responsible for strand cleavage at a given half-site. Does it come from the same recombinase molecule that provides the other active site residues for that half-site (*cis* cleavage) or is it from one of the remaining three subunits (*trans* cleavage)? The Cre–*loxA* structure presented here shows that Cre recombinase undergoes *cis* cleavage of the *loxA* site, in disagreement with a recent report³⁷. Lambda integrase and the Xer recombinases both cleave their recombination sites in *cis*^{38,39}; Flp recombinase cleaves in *trans*²⁷. The answer may lie both in the context-dependent nature of the C termini of the recombinases and in subtle structural differences among integrase family members. Based on the Cre–*loxA* complex, *trans* donation of helix M (which contains Tyr 324) to an adjacent active site is plausible: the helical segments are close together and only small changes in structure would be required (Fig. 2). This is particularly feasible for Flp recombinase, which contains a slightly longer linker between helices L and M⁴⁰; indeed, modelling of the Cre–*loxA* structure indicates that, given a longer linker, the M helix could be donated *in trans* and the remaining C-terminal residues could loop back to the same subunit to bury helix N *in cis*. These interactions would generate a 'trans-cyclic' arrangement in the synapse, a scheme that is similar to the 'asymmetric' mode proposed for Flp recombinase⁴¹.

In summary, our structure of Cre recombinase bound to a suicide DNA substrate provides a model for a λ integrase family site-specific

recombination synapse. The nearly four-fold symmetric tetramer of Cre–DNA half-sites formed by the association of two Cre–loxA complexes suggests a mechanism for recombination that does not involve branch migration of the Holliday junction intermediate between cleavage sites or isomerization of the protein–DNA architecture. Intersubunit contacts between C-terminal segments of Cre may regulate the ordered cleavage and strand-exchange steps of the reaction and these regions are likely to adopt conformations that depend on their molecular context. □

Methods

Crystallization of Cre–loxA. A Cre expression construct was prepared by polymerase chain reaction (PCR) amplification of the Cre coding region from bacteriophage P1 DNA and ligation into pET21a (Novagen). Both Cre and Se-Met Cre were purified from bacterial lysates on SP-Sepharose, Source-15S (Pharmacia) and CH15 hydroxyapatite (BioRad) columns. Crystals of the Cre–loxA complex were grown at 18°C by the hanging-drop method. Initial drops contained 4–8 µl 50 mM sodium acetate buffer, pH 5, 12% 2-methyl-2,4-pentanediol (MPD), 10 mM CaCl₂, 50 µM Cre, 75 µM loxA half-site, and reservoirs contained 100 mM acetate, pH 5, 24% MPD, 20 mM CaCl₂. Orthorhombic crystals (*a* = 107.7, *b* = 121.0, *c* = 180.4 Å; C222₁) grew in 3–4 days and reached a maximum size of 0.2 × 0.2 × 0.04 mm after 2–4 weeks.

Structure determination and refinement. Crystals were mounted in loops made of 20-µm nylon filaments and flash-frozen in liquid propane⁴². The asymmetric unit of this crystal form contained two Cre molecules and two loxA half-sites; the crystals diffracted well to 3 Å at a temperature of 100 K on laboratory sources and to at least 2.4 Å using synchrotron sources. Data were measured using a rotating anode X-ray source equipped with Charles Supper focusing mirrors and a MAR Research image-plate detector; NSLS X25 data were measured using a MAR Research image-plate detector, and NSLS X4A data were measured using Fuji image plates. All data were measured at 90–100 K and processed as described⁴². Two sets of dispersive differences (with respect to wavelength 1) and anomalous differences from a three-wavelength MAD experiment performed at NSLS X4A were treated as isomorphous and anomalous differences, respectively, and used together with the MIRAS data in a locally modified maximum-likelihood procedure⁴³ to phase the parent Cre/loxA structure. Modification of the solvent region⁴⁴ and phase combination with experimental probability distributions resulted in a clean electron density map with interpretable density for protein residues 21–340 and nearly all of the DNA bases. The two Cre molecules in the asymmetric unit were independently traced and non-crystallographic symmetry averaging was not used at any stage of the structure solution or refinement.

The Cre–loxA model was refined using simulated annealing coupled with torsion angle dynamics⁴⁵ at 3.0, 2.7 and 2.4 Å resolution, with model adjustments following each refinement round⁴⁶. A bulk solvent correction⁴⁷ allowed refinement with no low-resolution cutoff in the reflection data. A new parameter set was used to restrain nucleic acid geometry⁴⁸, with no torsional restraints imposed on the phosphotyrosine linkage. The final model includes 644 out of 686 amino acids (residues 20–341 in both Cre molecules), all nucleotides except one of the two overhanging 5' T residues, and 488 water molecules with unit occupancy and good hydrogen-bonding stereochemistry. All amino acids have (ϕ , ψ) backbone torsion angles in allowed regions of Ramachandran space.

Received 12 June; accepted 15 July 1997.

- Craig, N. L. The mechanism of conservative site-specific recombination. *Annu. Rev. Genet.* **22**, 77–105 (1988).
- Stark, W. M., Boocock, M. R. & Sherratt, D. J. Catalysis by site-specific recombinases. *Trends Genet.* **8**, 432–439 (1992).
- Kwon, H. J., Tirumalai, R., Landy, A. & Ellenberger, T. Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science* **276**, 126–131 (1997).
- Argos, P. *et al.* The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J.* **5**, 433–440 (1986).
- Abremski, K. E. & Hoess, R. H. Evidence for a second conserved arginine residue in the integrase family of recombination proteins. *Protein Eng.* **5**, 87–91 (1992).
- Landy, A. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu. Rev. Biochem.* **58**, 913–949 (1989).
- Abremski, K., Hoess, R. & Sternberg, N. Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. *Cell* **32**, 1301–1311 (1983).
- Abremski, K. & Hoess, R. Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *J. Biol. Chem.* **259**, 1509–1514 (1984).
- Sternberg, N., Hamilton, D., Austin, S., Yarmolinsky, M. & Hoess, R. Site-specific recombination and

- its role in the life cycle of bacteriophage P1. *Cold Spring Harbor Symp. Quant. Biol.* **1**, 297–309 (1981).
- Sauer, B. Manipulation of transgenes by site-specific recombination: use of Cre recombinase. *Meth. Enzymol.* **225**, 890–900 (1993).
- Kilby, N. J., Snaith, M. R. & Murray, J. A. Site-specific recombinases: tools for genome engineering. *Trends Genet.* **9**, 413–421 (1993).
- Tsurushita, N., Fu, H. & Warren, C. Phage display vectors for *in vivo* recombination of immunoglobulin heavy and light chain genes to make large combinatorial libraries. *Gene* **172**, 59–63 (1996).
- Qin, M., Bayley, C., Stockton, T. & Ow, D. W. Cre recombinase-mediated site-specific recombination between plant chromosomes. *Proc. Natl Acad. Sci. USA* **91**, 1706–1710 (1994).
- Lakso, M. *et al.* Targeted oncogene activation by site-specific recombination in transgenic mice. *Proc. Natl Acad. Sci. USA* **89**, 6232–6236 (1992).
- Betz, U. A., Voshchenko, C. A., Rajewsky, K. & Muller, W. Bypass of lethality with mosaic mice generated by Cre-loxP-mediated recombination. *Curr. Biol.* **6**, 1307–1316 (1996).
- Zou, Y. R., Muller, W., Gu, H. & Rajewsky, K. Cre-loxP-mediated gene replacement: a mouse strain producing humanized antibodies. *Curr. Biol.* **4**, 1099–1103 (1994).
- Kuhn, R., Schwenk, F., Aguet, M. & Rajewsky, K. Inducible gene targeting in mice. *Science* **269**, 1427–1429 (1995).
- Metzger, D., Clifford, J., Chiba, H. & Chambon, P. Conditional site-specific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase. *Proc. Natl Acad. Sci. USA* **92**, 6991–6995 (1995).
- Hickman, A. B., Waninger, S., Socca, J. J. & Dyda, F. Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HPI integrase at 2.7 Å resolution. *Cell* **89**, 227–237 (1997).
- Subramanya, H. S. *et al.* Crystal structure of the site-specific recombinase, XerD. *EMBO J.* **17**, 5178–5187 (1997).
- Schultz, S. C., Shields, G. C. & Steitz, T. A. Crystallization of *Escherichia coli* catabolite gene activator protein with its DNA-binding site. The use of modular DNA. *J. Mol. Biol.* **213**, 159–166 (1990).
- Pargellis, C. A., Nunes-Duby, S., de, V. L. & Landy, A. Suicide recombination substrates yield covalent lambda integrase–DNA complexes and lead to identification of the active site tyrosine. *J. Biol. Chem.* **263**, 7678–7685 (1988).
- Sigal, N. & Alberts, B. Genetic recombination: the nature of crossed strand-exchange between two homologous DNA molecules. *J. Mol. Biol.* **71**, 789–793 (1972).
- Duckett, D. R. *et al.* The structure of the Holliday junction, and its resolution. *Cell* **55**, 79–89 (1988).
- Hoess, R. H. & Abremski, K. Interaction of the bacteriophage P1 recombinase Cre with the recombining site loxP. *Proc. Natl Acad. Sci. USA* **81**, 1026–1029 (1984).
- Hoess, R., Abremski, K., Irwin, S., Kendall, M. & Mack, A. DNA specificity of the Cre recombinase resides in the 25 kDa carboxyl domain of the protein. *J. Mol. Biol.* **216**, 873–882 (1990).
- Chen, J. W., Lee, J. & Jayaram, M. DNA cleavage in trans by the active site tyrosine during Flp recombination: switching protein partners before exchanging strands. *Cell* **69**, 647–658 (1992).
- Rafferty, J. B. *et al.* Crystal structure of DNA recombination protein RuvA and a model for its binding to the Holliday junction. *Science* **274**, 415–420 (1996).
- Lavery, R. & Sklenar, H. The definition of generalised helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dynam.* **6**, 63–91 (1988).
- Hoess, R. H., Wierzbicki, A. & Abremski, K. The role of the loxP spacer region in P1 site-specific recombination. *Nucleic Acids Res.* **14**, 2287–2300 (1986).
- Stark, W. M., Sherratt, D. J. & Boocock, M. R. Site-specific recombination by Tn3 resolvase: topological changes in the forward and reverse reactions. *Cell* **58**, 779–790 (1989).
- Kitts, P. A. & Nash, H. A. Homology-dependent interactions in phage lambda site-specific recombination. *Nature* **329**, 346–348 (1987).
- Nunes-Duby, S., Azaro, M. A. & Landy, A. Swapping DNA strands and sensing homology without branch migration in lambda site-specific recombination. *Curr. Biol.* **5**, 139–148 (1995).
- Arciszewska, L. K., Grainge, I. & Sherratt, D. J. Action of site-specific recombinases XerC and XerD on tethered Holliday junctions. *EMBO J.* **16**, 3731–3743 (1997).
- Azaro, M. A. & Landy, A. The isometric preference of Holliday junctions influences resolution bias by lambda integrase. *EMBO J.* **16**, 3744–3755 (1997).
- Stark, W. M. & Boocock, M. R. Gatecrashers at the catalytic party. *Trends Genet.* **11**, 121–123 (1995).
- Shaikh, A. C. & Sadowski, P. D. The Cre recombinase cleaves the lox site in trans. *J. Biol. Chem.* **272**, 5695–5702 (1997).
- Nunes-Duby, S. *et al.* Lambda integrase cleaves DNA in cis. *EMBO J.* **13**, 4421–4430 (1994).
- Arciszewska, L. K. & Sherratt, D. J. Xer site-specific recombination in vitro. *EMBO J.* **14**, 2112–2120 (1995).
- Blakely, G. W. & Sherratt, D. J. Cis and trans in site-specific recombination. *Mol. Microbiol.* **20**, 233–238 (1996).
- Qian, X. H., Inman, R. B. & Cox, M. M. Protein-based asymmetry and protein–protein interactions in FLP recombinase-mediated site-specific recombination. *J. Biol. Chem.* **265**, 21779–21788 (1990).
- Van Duyne, G. D., Ghosh, S., Maas, W. K. & Sigler, P. B. Structure of the oligomerization and L-arginine binding domain of the arginine repressor of *Escherichia coli*. *J. Mol. Biol.* **256**, 377–391 (1996).
- Otwiński, Z. in *CCP4 Proc.* 80–88 (Daresbury Laboratory, Warrington, UK, 1991).
- Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
- Rice, L. M. & Brunger, A. T. Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Prot. Struct. Funct. Genet.* **19**, 277–290 (1996).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Jiang, J. S. & Brunger, A. T. Protein hydration observed by X-ray diffraction: solvation properties of penicillopepsin and neuraminidase crystal structures. *J. Mol. Biol.* **243**, 100–115 (1994).
- Parkinson, G., Vojtechovsky, J., Clowney, L., Brunger, A. T. & Berman, H. New parameters for the refinement of nucleic acid-containing structures. *Acta Crystallogr. D* **52**, 57–64 (1996).
- Kraulis, P. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Carson, M. Ribbons 2.0. *J. Appl. Crystallogr.* **24**, 958–961 (1991).

Acknowledgements. We thank L. Berman and Z. Yin for access to and help with the NSLS X25 beamline, W. Hendrickson and C. Ogata for access to and help with the NSLS X4A beamline, J. Ni for assistance with data collection, Y.-M. Zheng for technical support, R. Hoess for valuable discussions, and M. Lemmon, H. Nelson, M. Lewis, H. Lu and K. Ferguson for helpful comments. This work was partly supported by ACS IRG and NCI core support grants through the V. Penn. Cancer Center.

Correspondence and requests for materials should be addressed to G.D.V. (e-mail: vanduyne@crystal.med.upenn.edu). Coordinates have been deposited in the protein data bank at Brookhaven National Laboratory with accession code 1CRX.

Evidence for a downward mass flux in the penumbral region of a sunspot

C. Westendorp Plaza*, J. C. del Toro Iniesta*, B. Ruiz Cobo*, V. Martínez Pillet†, B. W. Lites† & A. Skumanich†

* Instituto de Astrofísica de Canarias, E-38200, La Laguna, Tenerife, Spain

† High Altitude Observatory, NCAR, Boulder, PO Box 300, Colorado 80307, USA

Sunspots were the first extraterrestrial phenomenon found to harbour magnetic fields^{1,2}. But the physical nature of sunspots and their relationship to the Sun's global magnetic field are still poorly understood³. Perhaps the largest uncertainty is related to the outermost region of sunspots (the penumbra) and, in particular, the nature of the so-called Evershed⁴ flow—a stream of material emanating radially from sunspots at velocities of up to $\sim 6 \text{ km s}^{-1}$ (ref. 5), before vanishing abruptly at the outer penumbral edges. Here we make use of a recently developed optical tomographic technique⁶ to obtain a three-dimensional model of the magnetic field and mass flow in the vicinity of a sunspot. We find that some of the magnetic field lines, together with a significant part of the Evershed mass flux, flow back towards the Sun in the deepest atmospheric layers at the outer edge of the sunspot and its surroundings. This observation should provide an important clue to our understanding of the appearance, stability and decay of sunspots, the most conspicuous tracers of the solar activity cycle.

The Evershed effect consists of a shift of the absorption line profiles towards shorter wavelengths (that is, blueshift) in the penumbra closest to the centre of the solar disk, and a redshift on the opposite side farthest from the disk centre. The effect produces asymmetric absorption line profiles, in the sense that the line wings are more shifted than the line core. This asymmetry in turn results in a net circular polarization when integrated over the full line profile, and has been linked to a gradient of the line-of-sight component of the velocity^{7–9}. The decrease in velocity with height has also been determined by simultaneously observing several spectral lines which probe different heights in the solar atmosphere. Spectral lines forming at very high levels (in the solar chromosphere) even show that the direction of the motion is an inflow rather than an outflow at these heights. The Evershed velocities disappear when the spot is close to the disk centre, and steadily grow as the spot approaches the solar limb, suggesting that the flows are nearly parallel to the solar surface at average photospheric levels.

Through the use of the inversion of the radiative transfer equation for polarized light⁶, we have obtained an optical tomography of a sunspot. This inversion technique is based on the so-called response functions which have proved to give the sensitivity of Stokes profiles to variations of physical quantities^{10,11}, allowing one to retrieve the stratification of such quantities with optical depth. The inherently ill-conditioned problem of every inversion is worked out by means of a modified singular-value-decomposition technique. In this way, information is selected from regions where response functions are significantly different from zero. When applied to two-dimensional maps of the full Stokes spectral profiles from a sunspot—as measured with the Advanced Stokes Polarimeter¹² operated at the VTT, National Solar Observatory—a complete three-dimensional model of the vector magnetic field, Doppler velocity, and other thermodynamic parameters results.

Two views of the fairly typical, round sunspot that we studied are shown in Fig. 1. (The spot belongs to the region NOAA 7197 as observed on 18 June 1992 around 16:30 UT at latitude -11.4° ,

longitude 8.8°W .) Figure 1a is a continuum image where the outer boundary of the spot (92% contrast) is marked by a dashed line. An angular resolution between 1 and 2 arcsec can be discerned. Figure 1b is a map from our inversion technique of the inclination of the magnetic field and Doppler velocity within the lowest atmospheric layer (continuum optical depth unity) to which visible spectral lines (in our case, the two Fe I lines at 630.15 and 630.25 nm) have significant sensitivity. The colour coding in Fig. 1b corresponds to the magnetic field inclination as measured from the normal to the solar surface. In this image we note the presence of regions all around the spot with an inclination $>90^\circ$ (blue patches): the magnetic field in these regions is therefore directed back into the Sun. This is an important result: although theoretical models¹³ and suggestions based on current knowledge^{14–16,29} have been proposed, very little magnetic return flux has been observed so far. In Fig. 1b the Evershed effect is indicated by the Doppler velocity signatures (solid lines): a blueshift (thick black lines, -3 km s^{-1} ; thin black lines, -1.5 km s^{-1}) in the penumbra towards disk centre (indicated by the arrow) and a redshift (white lines, $+3 \text{ km s}^{-1}$) towards the limb. Quite evidently, the Evershed effect does not stop at the visible boundary of the sunspot (dashed lines in Fig. 1a, b). What is clear is that the Evershed effect is highly correlated with the magnetic field inclination, as would be expected from the 'freezing' of plasma to magnetic field lines of force should this effect be indeed the signature of a mass flow. Blueshifts (upflows) are mostly located at the middle penumbra where the magnetic field points upward. More importantly, redshifts (downflows), at the outer penumbra, almost coincide with fields pointing downward into the surface.

Layers around continuum optical depth unity have traditionally been considered inaccessible by observing medium-strong lines like

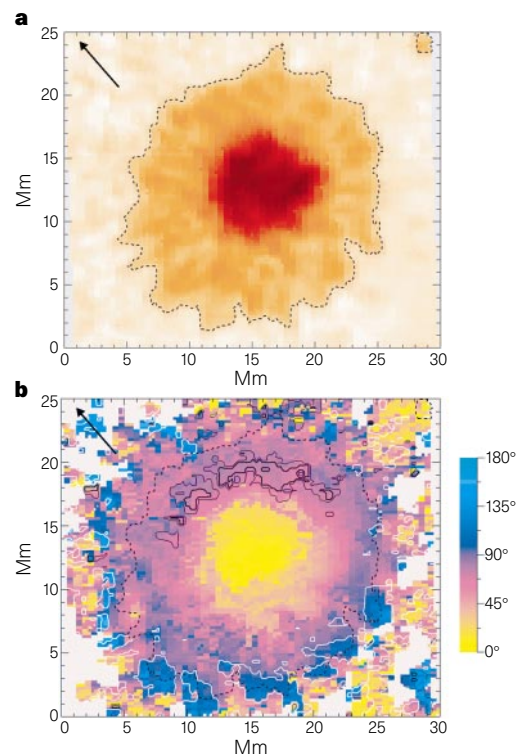


Figure 1. Simple round sunspot near the centre of the solar disk. Dashed lines show the visible limits of the spot. **a**, Conventional visible-light continuum image. **b**, Lowest observable layer obtained from the inversion of the data (see text). The magnetic field inclination as measured from the normal to the solar surface is shown in colour, the velocity towards (black lines) and away from (white lines) the observer as contours. (Thick black lines, -3 km s^{-1} ; thin black lines, -1.5 km s^{-1} ; white lines, $+3 \text{ km s}^{-1}$). The arrow points to the centre of the solar disk.

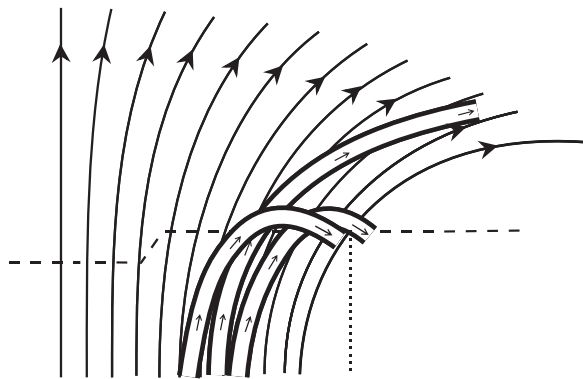


Figure 2 Diagram of penumbral structure as inferred from our results. Some magnetic tubes carry the Evershed flow. Most of the mass flux returns into the Sun through the low-lying loops. The remaining flux reaches the superpenumbral canopy. The dashed line represents the continuum optical depth unity surface (not to scale). The visible penumbral boundary is indicated by the vertical dotted line.

the ones we use here. However, in the presence of gradients of the line-of-sight velocity, symmetric wavelength samples (with respect to line centre) probe different layers^{11,14}. It is these gradients which make the response functions at low layers become significant, hence enabling accurate inferences where former techniques are not sensitive. The uncertainties of our retrieved values are inversely proportional to the magnitude of the response functions. At continuum optical depth unity, typical values of the errors in the line-of-sight velocity are of order of 0.5 km s^{-1} , 100 G in the magnetic field strength and 5° in the magnetic inclination. These error estimates are consistent with Monte Carlo simulation experiments (inaccuracies due to different initializations are negligible). At the same time, it is worth noting that our results do agree with techniques that do not acknowledge the presence of gradients at those higher layers which they sample.

We note that all $\sim 4,000$ points of the sunspot map have been inverted independently and a coherent picture of the spot has been obtained. The results retrieved in this way all around the spot confirms the reliability of the method. All points of our map showing the vector magnetic field and velocity pointing downwards share a common characteristic mostly visible in the profiles of circular polarization (Stokes V): a complex shape (with three lobes) unlike the usual, antisymmetric (two lobes) appearance. These characteristic profiles appear on both the limb- and the centre-side penumbra allowing us to discard possible projection effects. The presence of these spectral signatures in Stokes V can only be interpreted in terms of two polarities along the line-of-sight and/or averaged within our resolution element. Moreover, because of the redshifted nature of the extra third lobe, the polarity pointing into the Sun necessarily carries a downflow.

A diagram of the model we envisage from our results is shown in Fig. 2. Low-lying magnetic flux tubes actually carrying most of the Evershed flow re-enter the solar surface in the outermost parts of the penumbra and close by. Differential opacity effects¹⁷ between limb- and centre-side penumbra make this re-entering more evident in the limb-side part of the map in Fig. 1b. Our results at higher layers are compatible with the horizontal part of the loops shown in Fig. 2.

Other (more elevated) magnetic tubes extend well outside the sunspot limits (indicated by the vertical dotted line in Fig. 2). These tubes carry just a small portion of the flow and are identified with the already known magnetic “superpenumbral canopy”, where fairly horizontal fields overlie the non-magnetic atmosphere^{18,19}. Such sunspot canopies are known to host line-of-sight velocities²⁰ of up to 2 km s^{-1} . Just at the outer penumbral edge, we detect material

velocities at either very low layers (downward flowing) and very high layers (slightly upward), the latter being significantly smaller in magnitude. Such velocity variations are in agreement with the magnetic inclination gradients we find. This picture reconciles the long-lasting controversy based on the claimed abrupt end of the Evershed effect at the outer sunspot limits: although some authors affirm that this effect stops at the penumbral boundary^{21,22}, others detect it persisting farther out^{15,17,23}. Observations in line with both sides of this debate have been made at very high angular resolution^{17,21,22} (resolving structures smaller than 0.5 arcsec, about 350 km on the Sun), so the differing interpretations cannot be attributed to angular resolution; they must arise from the sensitivity of the various observing and analysis techniques to separate heights of the solar atmosphere.

The topology we infer can explain why the Evershed effect seems to vanish outside the spot while still conserving mass. At first glance, flows passing into the canopy might have provided an explanation for mass conservation. On closer examination however, the gas density at these heights is an order of magnitude smaller than in the penumbra, where the largest Evershed shifts are measured, so the canopy cannot host all of the Evershed flow of the penumbra^{20,24}. The inner penumbral flux is 14 times larger than the canopy flux²⁰, a situation clearly unsatisfactory from a mass conservation point of view. With our findings of field lines bending back down at the end of the penumbra, a better balance can be established. When we estimate the ratio of the vertical (downflowing) mass flux just outside the penumbral boundary to the vertical (upflowing) flux in the inner penumbra, values in the range of 1.2–2.2 are obtained. The different values inferred depend on the average with height that is used. It is important to point out that these estimates were calculated after assuming that the flows were directed along the magnetic field lines. The agreement between both mass fluxes is remarkable given the uncertainties in them. In particular, the inner and outer mass fluxes were computed at different angular sections within the penumbra. As the spot is 10° away from disk centre, the sensitivity of the observations to different penumbral flows depends appreciably on the angular section we choose. The two mass fluxes were estimated at the angular sector where their corresponding signatures were more conspicuous.

A greater challenge is to understand the physical mechanism responsible for such a flow. A prominent hypothesis for the Evershed effect is the “siphon flow” mechanism^{24,25,28,29}, in which differences in gas pressure drive a flow along the field lines within arched magnetic tubes. Compared to the inner footpoint (the deepest observable portion of the magnetic tube), the outer footpoint must have a higher magnetic field, hence a lower gas pressure to maintain pressure equilibrium with its surroundings, in order to generate an outward flow. This comparison of magnetic fields, however, has to be made at those heights in each footpoint that have the same gravitational potential. Unfortunately, the differential opacity between the inner and outer footpoints prevents such a comparison with current observational means.

Our global picture of an Evershed flow returning into the Sun has to be compatible with observations carried out at other disk positions. Preliminary results based on the analysis of another spot observed with the same instrument confirm this view.

The results reported here are based on the best current spectro-polarimetric data having moderate to high spatial resolution. But the penumbra has significant fine structure²⁶, analysis of which may reveal a possible sub-structure both horizontally and vertically. Therefore, observations with higher spatial resolution would be valuable. Future work should also consider time variations; we have presented a static ‘snapshot’ of the complex, dynamic penumbra²⁷, a full understanding of which is still to be accomplished. $\square$

Received 21 March; accepted 15 July 1997.

1. Hale, G. E. Solar vortices. *Nature* **78**, 368–369 (1908).
2. Hale, G. E. On the probable existence of a magnetic field in sun-spots. *Astrophys. J.* **28**, 315–343 (1908).

3. Livingston, W. Radial filamentary structure in a sunspot umbra. *Nature* **350**, 45–46 (1991).
4. Evershed, J. Radial movement in sun-spots. *Mon. Not. R. Astron. Soc.* **69**, 454–457 (1909).
5. Thomas, J. H. & Weiss, N. N. in *Sunspots: Theory and Observations* (eds Thomas, J. H. & Weiss, N. A.) 3–59 (Kluwer, Dordrecht, 1992).
6. Ruiz Cobo, B. & del Toro Iniesta, J. C. Inversion of Stokes profiles. *Astrophys. J.* **398**, 375–385 (1992).
7. Makita, M. & Kawakami, H. A study of line asymmetry in unipolar sunspots. *Publ. Astron. Soc. Jpn.* **38**, 257–265 (1996).
8. Sánchez Almeida, J. & Lites, B. W. Observation and interpretation of the asymmetric Stokes Q, U, and V line profiles in sunspots. *Astrophys. J.* **398**, 359–374 (1992).
9. Skumanich, A. & Lites, B. W. The polarization properties of model sunspots: the broad-band polarization signature of the Schlüter–Temesváry representation. *Astrophys. J.* **322**, 483–493 (1987).
10. Landi Degl’Innocenti, E. & Landi Degl’Innocenti, M. Response functions for magnetic lines. *Astron. Astrophys.* **56**, 111–115 (1987).
11. Ruiz Cobo, B. & del Toro Iniesta, J. C. On the sensitivity of Stokes profiles to physical quantities. *Astron. Astrophys.* **283**, 129–143 (1994).
12. Elmore, D. F. *et al.* The Advanced Stokes Polarimeter. A new instrument for solar magnetic field research. *Proc. SPIE* **1746**, 22–33 (1992).
13. Osherovich, V. A. A new magneto-hydrostatic theory of sunspots. *Solar Phys.* **77**, 63–68 (1982).
14. del Toro Iniesta, J. C., Tarbell, T. D. & Ruiz Cobo, B. On the temperature and velocity through the photosphere of a sunspot penumbra. *Astrophys. J.* **436**, 400–410 (1994).
15. Stanchfield, D. C. H., Thomas, J. H. & Lites, B. W. The vector magnetic field, Evershed flow, and intensity in a sunspot. *Astrophys. J.* **477**, 485–494 (1997).
16. Maltby, P. in *1st Advances in Solar Physics Euroconf. Advances in the Physics of Sunspots* (eds Schmieder, B., del Toro Iniesta, J. C. & Vázquez, M.) 91–110 (Vol. 118, ASP Conf. Ser., Astron. Soc. Pac., San Francisco, 1997).
17. Rimmele, T. R. Evidence for thin elevated Evershed channels. *Astron. Astrophys.* **298**, 260–276 (1995).
18. Giovanelli, R. & Jones, H. P. Three-dimensional structure of atmospheric magnetic fields. *Solar Phys.* **79**, 267–278 (1982).
19. Solanki, S. K., Rüedi, I. & Livingston, W. Infrared lines as probes of solar magnetic features. V. The magnetic structure of a simple sunspot and its canopy. *Astron. Astrophys.* **263**, 339–350 (1992).
20. Solanki, S. K., Montavon, C. A. P. & Livingston, W. Infrared lines as probes of solar magnetic features. VII. On the nature of the Evershed effect in sunspots. *Astron. Astrophys.* **283**, 221–231 (1994).
21. Title, A. M. *et al.* On the magnetic and velocity field geometry of simple sunspots. *Astrophys. J.* **403**, 780–796 (1993).
22. Wiehr, E. & Degenhardt, D. The Evershed effect in penumbral fine structures. II. Spatial correlation analysis. *Astron. Astrophys.* **287**, 625–632 (1994).
23. Rimmele, T. R. Sun center observations of the Evershed effect. *Astrophys. J.* **445**, 511–516 (1995).
24. Thomas, J. H. in *Solar Surface Magnetism* (eds Rutten, R. J. & Schrijver, C. J.) 219–235 (Kluwer, Dordrecht, 1994).
25. Mayer, F. & Schmidt, H. U. Magnetisch ausgerichtete Strömungen zwischen Sonnenflecken. *Z. Angew. Math. Mech.* **48**, 218–221 (1968).
26. Lites, B. W. Physics of sunspots: dynamics and fine structure. *IAU Trans. A* (in the press).
27. Shine, R. A. *et al.* High-resolution observations of the Evershed effect in sunspots. *Astrophys. J.* **430**, 413–424 (1994).
28. Thomas, J. H. & Montesinos, B. A siphon-flow model of the photospheric Evershed flow in a sunspot. *Astrophys. J.* **407**, 398–401 (1993).
29. Thomas, J. H. in *Solar and Astrophysical Magnetohydrodynamic Flows* (ed. Tsinganos, K. C.) 39–60 (Kluwer, Dordrecht, 1996).

Acknowledgements. This work was supported in part by the Spanish DGES. The NCAR is sponsored by the National Science Foundation.

Correspondence and requests for materials should be addressed to C.W.P. (e-mail: westend@lliac.es)

A ductile ceramic eutectic composite with high strength at 1,873 K

Y. Waku, N. Nakagawa, T. Wakamoto, H. Ohtsubo, K. Shimizu & Y. Kohtoku

Ube Research Laboratory, Corporate Research & Development, UBE Industries, Ltd, Ube City, Yamaguchi, 755, Japan

Monolithic ceramics are not widely used as structural materials because of their brittleness. Ceramic matrix composites, in which whiskers^{1–3} or fibres^{4–7} of strong ceramics such as silicon carbide or silicon nitride are embedded in a ceramic matrix, offer improved toughness and strength because the energy of cracks may be dissipated at the whisker/matrix interface. Here we describe a strong ceramic composite with a different kind of microstructure, made by unidirectional solidification of an Al₂O₃/GdAlO₃ eutectic mixture. This composite has a microstructure in which continuous networks of single-crystal Al₂O₃ and single-crystal GdAlO₃ interpenetrate without grain boundaries. Rather than brittle fracture, the material displays plastic deformation at 1,873 K owing to dislocation motion, as in metals. The high

strength and resistance to brittle failure of this material at such high temperatures augurs well for applications in mechanical engineering.

Microstructural studies on an Al₂O₃/Y₃Al₅O₁₂ (YAG) system using the Bridgman method show that the microstructure of this composite can be controlled by unidirectional solidification⁸. It has been reported that a unidirectionally solidified Al₂O₃/YAG eutectic composite has superior flexural strength, thermal stability and creep resistance at high temperature^{9–11}, and is a candidate for a high-temperature structural material. However, this material contains many grain boundaries or colonies between Al₂O₃ and YAG, which are expected¹² to impair the mechanical properties. Waku *et al.*^{13–15} have recently fabricated a eutectic composite consisting of single-crystal Al₂O₃ and single-crystal YAG with neither colonies nor grain boundaries, using a unidirectional solidification. The composite is thermally stable and its flexural strength at room temperature can be maintained up to 2,073 K, which is just below the melting point (~2,100 K). In addition, the compression creep strength at 1,873 K and a strain rate of 10^{–4} s^{–1} is ~13 times higher than that of sintered composites with the same composition, and the material shows neither weight gain nor grain growth even on heating at 1,973 K in air for 250 hours.

Slow directional solidification of the melt (5–150 mm h^{–1}) of a Ni–30Al (atom%) intermetallic compound results in an aligned rod-like γ' microstructure in a single-crystal NiAl matrix¹⁶. For the same volume fraction of reinforcement, the increase in ductility and toughness are dependent on the morphology of the second phase, with continuous microstructures offering the largest gains in improved ductility¹⁶. At the slowest solidification rate of 5 mm h^{–1} the rods are essentially continuous but discrete.

We have now attempted to fabricate a unidirectionally solidified Al₂O₃/GdAlO₃ composite with high strength and high ductility at high temperature by controlling the microstructure using a eutectic reaction of the Al₂O₃/Gd₂O₃ binary system. Commercially available Al₂O₃ powder (AKP-30, Sumitomo Chemical Co. Ltd., Tokyo, Japan) and Gd₂O₃ powder (Gd₂O₃-RU, Shin-etsu Chemical Co. Ltd., Tokyo, Japan) were mixed to a mole ratio of Al₂O₃/Gd₂O₃ = 77/23, and wet ball-milling using ethanol was carried out to obtain a homogeneous mixed powder. The slurry obtained was dried in a rotary evaporator to remove the ethanol. After pre-forming the mixed powder into a pellet 10 mm in diameter and 10 mm in height, it was premelted in an arc discharge. All unidirectional solidification experiments were carried out by using the advanced-alloy crystalline-structure-controlling equipment at the Japan Ultra-high Temperature Materials Research Center. The powder obtained by crushing the arc-melted ingots was put in a molybdenum crucible and placed in a chamber. The melting experiment was performed in the Mo crucible heated by high-frequency induction heating under a pressure of 10^{–5} mm Hg of argon, and then, after holding for 30 min at 2,123 K, unidirectional solidification was carried out by lowering the Mo crucible at a speed of 5 mm h^{–1} in the same argon atmosphere. For accurate control of crystal growth, a mini-crucible (2 mm in diameter and 17 mm in length) for growing a suitable seed crystal was set at the bottom of the main crucible.

Some of the mixed powder (Al₂O₃ + Gd₂O₃) was not treated as described above, but was hot-pressed in a carbon die (for 1 h at 1,973 K in vacuum (10^{–2} mm Hg)) to fabricate a sintered Al₂O₃/GdAlO₃ composite of dimensions 50 × 60 × 5 mm. This sintered composite had the same composition as the eutectic composite.

Three-point flexural tests were carried out by using specimens (3 × 4 × 36 mm) having the long axis parallel to the solidification direction. The equipment used in this study is the high-temperature uniaxial tension, compression and bending test system (a modified creep and fatigue machine, Instron type 8562). Flexural tests were conducted in an argon gas atmosphere at 1,873 K. The crosshead speed was 0.5 mm min^{–1}. Room-temperature fracture toughness

was measured by the indentation method using the equation of Marshall and Evans¹⁷:

$$K_{IC} = 0.0036E^{0.4}P^{0.6}a^{-0.7}(c/a)^{-1.5}$$

where E is the elastic modulus (373 GPa for this composite), P is the applied load, and a and c are characteristic dimensions of the Vickers impression and the radial/median crack, respectively. The surface was polished for Vickers indentation, and a load of 98 MPa was used for 15 s to induce indentation. Structural analyses were undertaken using a Rigaku-Denki RAD-RB X-ray diffraction apparatus. The transmission electron microscope (TEM) observations of the interface and grain-boundary structure between the phases used a Japan Electron JEM-2010, electron probe microanalysis (EPMA) analyses were undertaken using a Japan Electron JMX-8621MX.

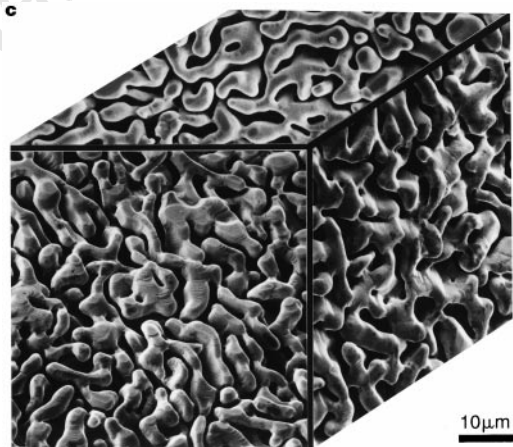
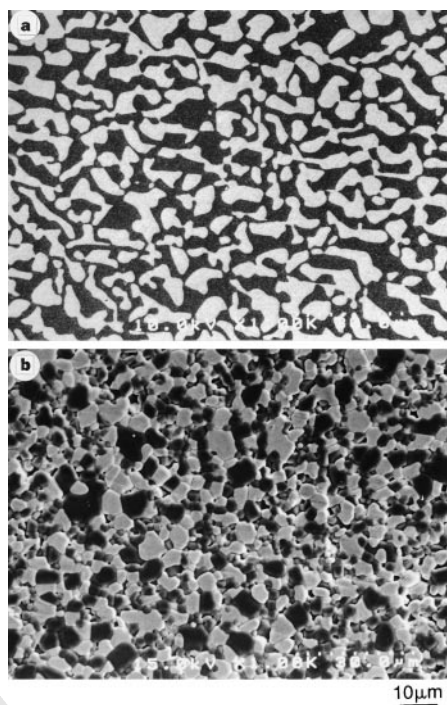


Figure 1 **a**, SEM image showing the microstructure of a cross-section perpendicular to the solidification direction of the unidirectionally solidified $\text{Al}_2\text{O}_3/\text{GAP}$ eutectic composite. **b**, SEM image showing the microstructure of a cross-section parallel to the hot-pressed plane of the sintered $\text{Al}_2\text{O}_3/\text{GAP}$ composite. **c**, SEM micrographs showing the three-dimensional configuration of single-crystal GAP in unidirectionally solidified $\text{Al}_2\text{O}_3/\text{GAP}$ eutectic composite.

The microstructures of the upper, middle and lower planes perpendicular to the solidification direction of the unidirectionally solidified eutectic composite, and those of the hot-pressed plane of the sintered composite with the same composition, consisted of Al_2O_3 and GdAlO_3 (GAP: gadolinium aluminium perovskite) phases; these were determined from X-ray diffraction patterns. Figure 1a and b show scanning electron microscopy (SEM) images of the cross-section perpendicular to the solidification direction of the eutectic composite and parallel to the hot-pressed plane of the sintered composite. The light area in the SEM micrograph is the GAP phase, and the dark area is the Al_2O_3 phase (identified by EPMA analysis). The dimensions of the microstructures are similar for both composites. Homogeneous microstructures with no pores or colonies are observed in the eutectic composite (Fig. 1a). In the X-ray diffraction pattern for the unidirectionally solidified eutectic composite, diffraction peaks from the (214) plane of the Al_2O_3 phase and from the (111) plane of the GAP phase are observed only in the plane inclined by 7° from the plane perpendicular to the solidification direction. Consequently, it can be concluded that this composite consists of (214) single-crystal Al_2O_3 with a hexagonal structure and (111) single-crystal GAP with a perovskite structure. In contrast, for the sintered composite, diffraction peaks from various planes are observed, characteristic of a polycrystalline ceramic composite with random crystal orientations. Figure 1c shows an SEM micrograph which illustrates the three-dimensional configuration of the single-crystal GAP in the unidirectionally solidified eutectic composite from which Al_2O_3 phases had been removed by heating in graphite powder at 1,923 K for 2 h. The configuration of the single-crystal GAP is a three-dimensionally connected porous structure, of irregular shape (Fig. 1c). We therefore conclude that the present unidirectionally solidified composite has a microstructure consisting of three-dimensionally continuous and complexly entangled single-crystal Al_2O_3 and single-crystal GAP. The microstructure was fabricated by controlling accurately the crystal growth and solidification process in the unidirectional solidification. This composite's room-temperature fracture toughness estimated from the indentation method is around $5\text{--}6\text{ MPa m}^{1/2}$ —a little higher than the value of $4\text{--}5\text{ MPa m}^{1/2}$ for an advanced Si_3N_4 ceramic¹⁸.

The microstructure of the $\text{Al}_2\text{O}_3/\text{GAP}$ eutectic composite consists of three-dimensionally bicontinuous single crystals. Analogous microstructures are observed in 'soft' systems, such as liquid

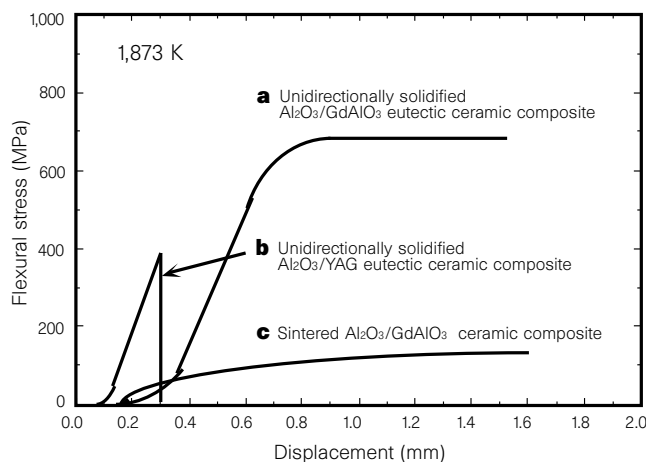


Figure 2 Typical stress-displacement curves in the three-point flexural test at 1,873 K of unidirectionally solidified $\text{Al}_2\text{O}_3/\text{GAP}$ eutectic composites (trace a) compared with unidirectionally solidified $\text{Al}_2\text{O}_3/\text{YAG}$ eutectic composites (trace b) and sintered $\text{Al}_2\text{O}_3/\text{GAP}$ composites (trace c).

crystals, surfactant–water and water–oil–soap mixtures and copolymers^{19,20}. These microstructures can show some similarity to ordered periodic minimal surfaces²⁰, but those in our composite are more analogous to the disordered ‘sponge’ phase of surfactants in water.

Figure 2 shows typical stress–displacement curves of the unidirectionally solidified eutectic composite and sintered composite obtained from the flexural test at 1,873 K. For comparison, a stress–displacement curve for a $\text{Al}_2\text{O}_3/\text{YAG}$ eutectic composite¹⁴ is also shown. The $\text{Al}_2\text{O}_3/\text{GAP}$ eutectic composite shows yielding behaviour under high stress, with a flexural yield stress of ~ 695 MPa. The flexural yield stresses for the other two specimens of this type are 705 and 670 MPa, and the average yield stress is

690 MPa. Furthermore, the relative fracture energy of the $\text{Al}_2\text{O}_3/\text{GAP}$ eutectic composite (which can be inferred from the area under the stress–strain curve) is dozens of times that of the $\text{Al}_2\text{O}_3/\text{YAG}$ eutectic composite, which undergoes brittle fracture at this temperature. By contrast, the sintered $\text{Al}_2\text{O}_3/\text{GAP}$ composite undergoes plastic deformation at lower stress, with less fracture energy. The $\text{Al}_2\text{O}_3/\text{GAP}$ eutectic composite did not fracture even after the flexural test, showing its substantial plasticity.

The existence of amorphous phases at interfaces or grain boundaries generally leads to a reduction in the strength of the material at high temperature^{21,22}. Figure 3 shows typical high-resolution TEM images of the boundary between Al_2O_3 and GAP phases in the sintered composite, and in the unidirectionally solidified eutectic

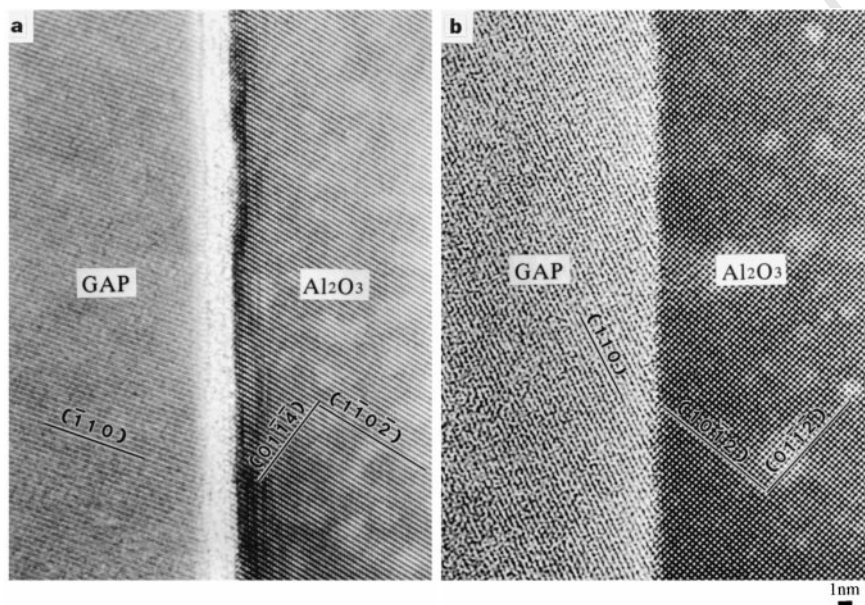


Figure 3 High-resolution TEM images of the grain boundaries between Al_2O_3 and GAP phases of sintered composites (**a**) and the interface between Al_2O_3 and GAP phases of the unidirectionally solidified $\text{Al}_2\text{O}_3/\text{GAP}$ eutectic composites (**b**). The beam directions are $[82\bar{1}03]\text{Al}_2\text{O}_3$ for **a** and $[2201]\text{Al}_2\text{O}_3$ for **b**.

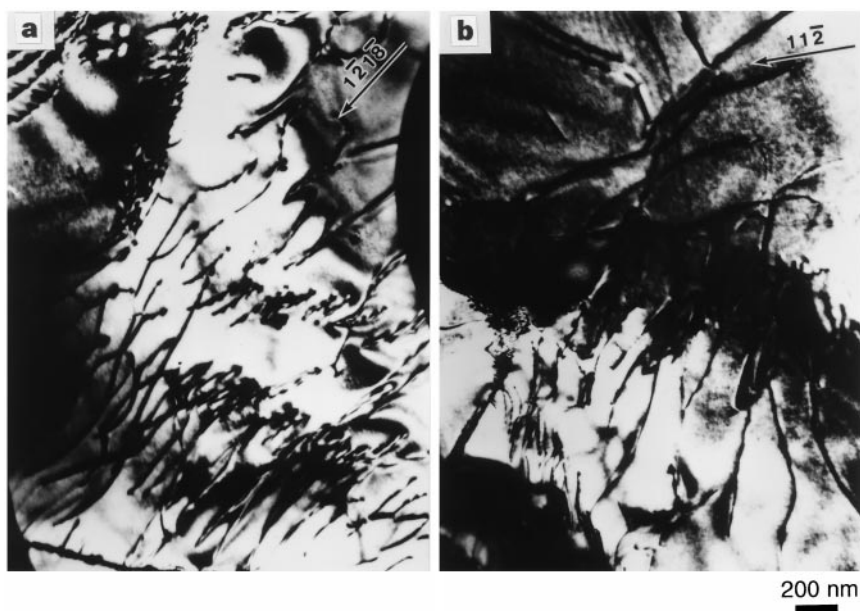


Figure 4 TEM images showing the dislocation structure of Al_2O_3 phases (**a**) and GAP phases (**b**) at the tensile-stressed side of the plastically deformed specimen after the three-point flexural test at 1,873 K of the unidirectionally solidified $\text{Al}_2\text{O}_3/\text{GAP}$

eutectic composites. The beam direction is nearly along the $[\bar{1}082\bar{3}]$ axis with the operating $\mathbf{g} = 1218$ for Al_2O_3 , and along the $[\bar{3}1\bar{1}]$ axis with the operating $\mathbf{g} = 112$ for GAP ($\mathbf{g}$ is the reciprocal lattice vector).

composite. Amorphous phases are observed at the grain boundary of the sintered composite (Fig. 3a), and also at the triple-grain junctions, grain boundaries between Al_2O_3 phases and between GAP phases in the sintered composite; but for the eutectic composite, no amorphous phases are observed at the interfaces between the Al_2O_3 and GAP phases and relatively compatible interfaces are formed (Fig. 3b).

Figure 4 shows TEM images of the dislocation structure observed at the tensile-stressed side of the plastically deformed specimen in the flexural test at 1,873 K for the unidirectionally solidified Al_2O_3 /GAP eutectic composites. Dislocations are observed in both single-crystal Al_2O_3 and GAP, showing that the plastic deformation occurred by dislocation motion. This indicates that the plastic deformation mechanism of the present eutectic composite is essentially different from that of micrograin superplasticity of ceramics²³ due to grain-boundary sliding or a liquid phase present at grain boundaries at high temperature.

We conclude that the superior high-temperature strength of Al_2O_3 /GAP eutectic composites over Al_2O_3 /YAG eutectic composites was obtained by the following means: a new combination of single-crystal Al_2O_3 with a hexagonal structure and single-crystal GAP with a perovskite structure; a microstructure consisting of three-dimensionally continuous and complexly entangled single-crystal Al_2O_3 and single-crystal GAP; a characteristic dimension of the microstructure of Al_2O_3 /GAP eutectic composites of around 3–5 μm (this dimension is defined as the typical length of the short axis of each domain seen in the cross-section perpendicular to the solidification direction), finer than that of ~20–30 μm for the Al_2O_3 /YAG eutectic composites; and the fact that no amorphous phase is formed at the interface between the Al_2O_3 and GAP phases (it is possible that the two phases are bound together strongly at this interface).

The maximum operating temperature of this ceramic eutectic composite can be expected to be ~1,973 K—much higher than that (1,323–1,373 K) of Ni-based single-crystal cast superalloys^{24,25} and that (~1,300 K) of oxide ceramics^{26,27}. Therefore, several useful applications can be considered, for example, gas turbines and power-generation systems with non-cooled turbine blades at very high temperatures. □

Received 10 February; accepted 7 July 1997.

1. Becher, P. F. & Wei, G. C. Toughening behavior in SiC-whisker-reinforced alumina. *J. Am. Ceram. Soc.* **67**, C267–C269 (1984).
2. Lio, S., Watanabe, M., Matsubara, M. & Matsuo, Y. Mechanical properties of alumina/silicon carbide whisker composites. *J. Am. Ceram. Soc.* **72**, 1880–1884 (1989).
3. Tamari, N., Tanaka, T., Kondou, I. & Kose, S. Sintering of alumina–silicon nitride whisker composites and their properties. *J. Ceram. Soc. Jpn* **99**, 370–375 (1991).
4. Takeda, M. et al. Properties of the low-oxygen-content silicon carbide fiber after high temperature heat treatment. *Ceram. Eng. Sci. Proc.* **12**, 1007–1018 (1991).
5. Ishikawa, T., Kajii, S., Matsunaga, K., Hogami, T. & Kohtoku, Y. Structure and properties of Si-Ti-C-O fiber-bonded ceramic material. *J. Mater. Sci.* **30**, 6218–6222 (1995).
6. Lamicq, P., Bernhart, G. A., Dauchier, M. M. & Mace, J. G. SiC/SiC composite ceramics. *Am. Ceram. Soc. Bull.* **65**, 336–338 (1986).
7. Lamicq, P. J. *Proc. Japan-Europe Symp. on Composite Materials* (ed. R&D Inst. of Metals and Composites for Future Industries) 4–9 (Japan Industrial Technology Assoc., Nagoya, 1993).
8. Viehnicki, D. & Schmid, F. Eutectic solidification in the system $\text{Al}_2\text{O}_3/\text{Y}_3\text{Al}_5\text{O}_{12}$. *J. Mater. Sci.* **4**, 84–88 (1969).
9. Mah, T. & Parthasarathy, T. A. Processing and mechanical properties of $\text{Al}_2\text{O}_3/\text{Y}_3\text{Al}_5\text{O}_{12}$ (YAG) eutectic composite. *Ceram. Eng. Sci. Proc.* **11**, 1617–1627 (1990).
10. Parthasarathy, T. A., Mah, T. & Matson, L. E. Creep behavior of an Al_2O_3 - $\text{Y}_3\text{Al}_5\text{O}_{12}$ eutectic composite. *Ceram. Eng. Sci. Proc.* **11**, 1628–1638 (1990).
11. Parthasarathy, T. A., Mah, T. & Matson, L. E. Deformation behavior of an Al_2O_3 - $\text{Y}_3\text{Al}_5\text{O}_{12}$ eutectic composite in comparison with sapphire and YAG. *J. Am. Ceram. Soc.* **76**, 29–32 (1993).
12. Stubican, V. S., Bradt, R. C., Kennard, F. L., Minford, W. J. & Sorrell C. C. in *Tailoring Multiphase and Composite Ceramics* (eds Tressler, R. E., Messing, G. L., Patano, C. G. & Newnham, R. E.) 103–114 (Mater. Sci. Res. Vol. 20, Plenum, New York, 1986).
13. Waku, Y., Nakagawa, N., Ohtsubo, H., Ohsora, Y. & Kohtoku, Y. High temperature properties of unidirectionally solidified Al_2O_3 /YAG composites. *J. Jpn. Inst. Metals* **59**, 71–78 (1995).
14. Waku, Y., Otsubo, H., Nakagawa, N. & Kohtoku, Y. Sapphire matrix composites reinforced with single crystal YAG phases. *J. Mater. Sci.* **31**, 4663–4670 (1996).
15. Waku, Y. et al. in *Processing and Fabrication of Advanced Materials* (eds Srivatsan, T. S. & Moore, J. J.) 323–339 (TMS, Cleveland, 1995).
16. Naebu, R. D., Misra, A. & Gibara, R. Plastic flow and fracture of B2 NiAl-based intermetallic alloys containing a ductile second phase. *Iron Steel Inst. Jpn* **31**, 1172–1185 (1991).
17. Marshal, D. B. & Evans, A. G. Reply to 'Comment on 'Elastic/Plastic Indentation Damage in Ceramics: The Median/Radial Crack System''. *J. Am. Ceram. Soc.* **64**, C182–C183 (1981).
18. Bar-On, I., Baratta, F. I. & Cho, K. Crack stability and its effect on fracture toughness of hot-pressed silicon nitride beam specimens. *J. Am. Ceram. Soc.* **79**, 2300–2308 (1996).

19. Antonietti, M. & Göltner, C. Superstructures of functional colloids: chemistry on the nanometer scale. *Angew. Chem. Int. Edn Engl.* **36**, 910–928 (1997).
20. Hyde, S. et al. *The Language of Shape—The Role of Curvature in Condensed Matter: Physics, Chemistry and Biology* (Elsevier, Amsterdam, 1997).
21. Clarke, D. R. High-resolution techniques and application to nonoxide ceramics. *J. Am. Ceram. Soc.* **52**, 236–246 (1979).
22. Echigoya, J., Hayashi, S., Sasaki, K. & Suto, H. Microstructure of directionally solidified MgO-ZrO₂ eutectic. *J. Jpn. Inst. Metals* **48**, 430–434 (1984).
23. Wakai, F. et al. A superplastic covalent crystal composite. *Nature* **344**, 421–423 (1990).
24. Goulette, M. J. in *Proc. 8th Int. Symp. on superalloys* (eds Kissinger, R. D. et al.) 3–6 (TMS, Pennsylvania, 1996).
25. Erickson, G. L. in *Proc. 8th Int. Symp. on superalloys* (eds Kissinger, R. D. et al.) 35–43 (TMS, Pennsylvania, 1996).
26. Davidge, R. W. & Evans, A. G. The strength of ceramics. *Mater. Sci. Eng.* **6**, 281–298 (1970).
27. Niihara, K. New design concept of structural ceramics—ceramic nanocomposites. *J. Ceram. Soc. Jpn* **99**, 974–982 (1991).

Acknowledgements. We thank M. Suzuki for his assistance in the TEM observations.

Correspondence and requests for materials should be addressed to Y.W. (e-mail: 25222u@ube-ind.co.jp).

Oxidative acylation using thioacids

Rihe Liu & Leslie E. Orgel

The Salk Institute for Biological Studies, PO Box 85800, San Diego, California 92186, USA

Several important prebiotic reactions, including the coupling of amino acids into polypeptides by the formation of amide linkages, involve acylation. These reactions present a challenge to the understanding of prebiotic synthesis. Condensation reactions relying on dehydrating agents are either inefficient in aqueous solution^{1,2} or require strongly acidic conditions and high temperatures³. Activated amino acids such as thioester derivatives have therefore been suggested as likely substrates for prebiotic peptide synthesis^{4–7}. Here we propose a closely related route to amide bond formation involving oxidative acylation by thioacids. We find that phenylalanine, leucine and phenylphosphate are acylated efficiently in aqueous solution by thioacetic acid and an oxidizing agent. From a prebiotic point of view, oxidative acylation has the advantage of proceeding efficiently in solution and under mild conditions. We anticipate that oxidative acylation should prove to be a general method for activating carboxylic acids, including amino acids.

The use of thioacids as acylating agents was studied extensively by Wieland and his co-workers. They found that thioacids do not acylate amines directly, although α -amino thioacids form peptides in aqueous solution via *N*-carboxyanhydride intermediates if the bicarbonate ion is present⁸. The only reference that we have found to the oxidative activation of thioacids as acylating agents is a brief remark by Sheehan and his co-workers who found that the oxidation of phthaloylthioglycine with glycine methyl ester in the presence of iodine–potassium iodide and excess of sodium bicarbonate leads to the precipitation of the peptide derivative⁹. We find that the oxidation of simple thioacids in aqueous solution generates surprisingly efficient acylating agents.

We first prepared a solution containing 25 mM phenylalanine and 75 mM thioacetic acid in HEPES buffer at pH 7.4 and added a two-fold excess of potassium ferricyanide. In a control reaction we omitted the ferricyanide and replaced air by argon. After 20 min the phenylalanine had been converted quantitatively (>99.5%) to *N*-acetyl phenylalanine (Fig. 1A), which was identified by fast atom bombardment mass spectroscopy (calculated for $\text{C}_{11}\text{H}_{13}\text{NO}_3\text{Na}$ ($\text{M}^+ + \text{Na}$), m/z 230.2; found, 230.1 ($\text{M}^+ + \text{Na}$, 100%)). In the control reaction <1% of product was formed (Fig. 1B). In a similar way we found that the acetylation of 25 mM leucine by 75 mM thioacetic acid in the presence of 150 mM potassium ferricyanide

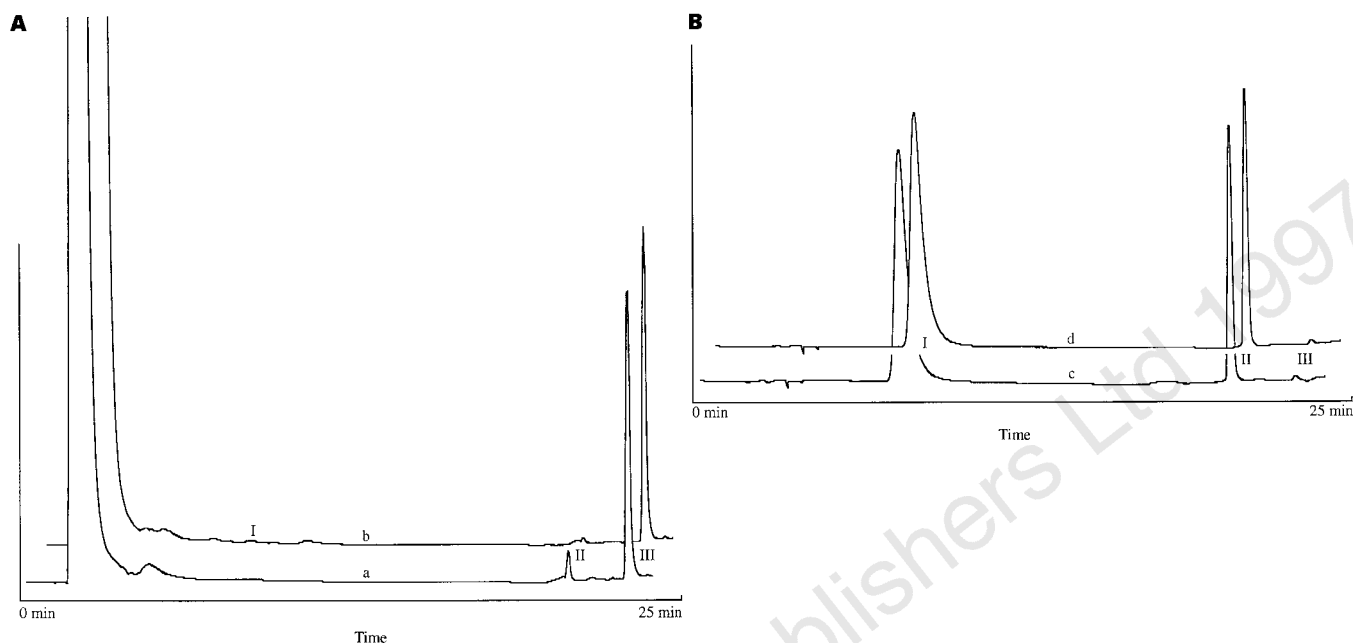


Figure 1 High-performance liquid chromatography (HPLC) elution profiles of the product from the oxidative acylation of phenylalanine using thioacetic acid. The reactions were carried out in 1.7 ml Eppendorf tubes in volumes of 200 μ l. The reaction mixtures contained 25 mM phenylalanine, 75 mM thioacetic acid and 150 mM $K_3Fe(CN)_6$ in HEPES buffer at pH 7.4. In control reactions the oxidizing agent was omitted and air was removed by passing argon through the solution. The reaction mixtures were incubated at room temperature. After appropriate times, an aliquot of the reaction mixture was taken, diluted with 200 μ l initial buffer

(0.1% TFA), and injected onto a reverse-phase C18 column. The products were eluted from the column with a linear gradient of 0–90% acetonitrile (0.1% TFA) at a flow rate of 1.0 ml min⁻¹ for 25 min. The eluate was monitored at 250 nm. **A**, Experiments with ferricyanide. Trace a, after 10 min; trace b, after 20 min. **B**, Experiments without ferricyanide. Trace c, at zero time; trace d, after 18 h. In **A** and **B**, peak I is thioacetic acid, peak II is phenylalanine and peak III is acetylphenylalanine.

led to quantitative (>99%) production of *N*-acetyl leucine after 20 min (data not shown).

When only one equivalent of thioacetic acid was used with ferricyanide as oxidant, ~50% of the phenylalanine was acylated rapidly; further acylation proceeded slowly. This suggests the sequence of reactions shown in Fig. 2. Clearly, $CH_3-CO-SS^-$ is a much better leaving group than HSS^- . When excess oxidant is present, the $CH_3-CO-SSH$ intermediate may be further oxidized to $CH_3-CO-S_4-COCH_3$, and so on. Comparison of acylation by ethylthioacetate with oxidative acylation by thioacetic acid showed that the latter is at least 100 times faster (data not shown).

Although the experiments described above suggest very strongly that the mechanism outlined in Fig. 2 is correct, they do not establish the mechanism conclusively. To provide further evidence for an oxidative mechanism, we repeated the acylation reaction, replacing ferricyanide by KI_3 or the Fe^{2+} ion in the presence of air. The yield of *N*-acetylphenylalanine is very similar with the three oxidizing agents (Fig. 3). As $[Fe(CN)_6]^{3+}$, KI_3 and Fe^{2+}/O_2 have little in common except their ability to oxidize substrates such as thiocompounds, we may safely conclude that the acylation of phenylalanine by thioacetic acid in the presence of ferricyanide is indeed dependent on oxidation.

We next studied the acylation of phenylphosphate, a substrate very different from an amino acid. The oxidation of 0.2 M thioacetic acid in the presence of 25 mM phenylphosphate yields acetylphenylphosphate in >40% yield (Fig. 4). The product was identified by fast atom bombardment mass spectroscopy (calculated for $C_8H_9PO_5Na$ ($M^+ + Na$), m/z 239.1; found, 238.7 ($M^+ + Na$, 100%)). The successful acylation of phenylalanine and phenylphosphate in aqueous solution suggests that the acylating agent generated by the oxidation of thioacetic acid is both powerful and general.

The use of thioesters in ligation reactions is the subject of several

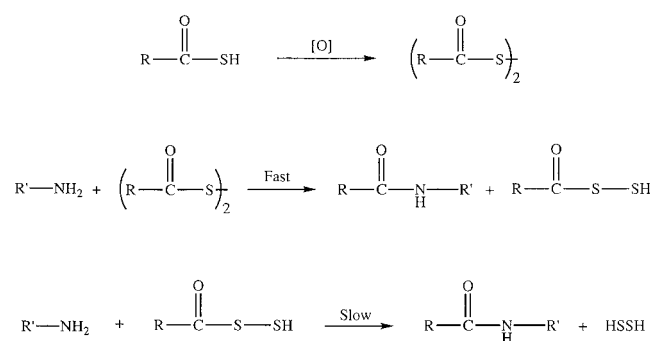


Figure 2 Oxidative acylation using thioacids.

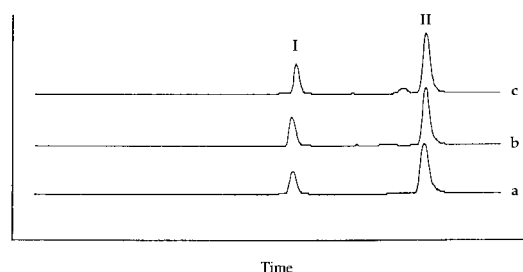


Figure 3 HPLC elution profile of products from the oxidative acylation of 25 mM phenylalanine using 25 mM thioacetic acid in the presence of 50 mM oxidizing agents; trace a, ferricyanide; trace b, KI_3 ; trace c, Fe^{2+} and air. The reactions were carried out in HEPES buffer at pH 7. The reaction time was 6 h. For details of analytical procedures see Fig. 1 legend. Peak I (21.4 min), phenylalanine; peak II (23.8 min), *N*-acetylphenylalanine.

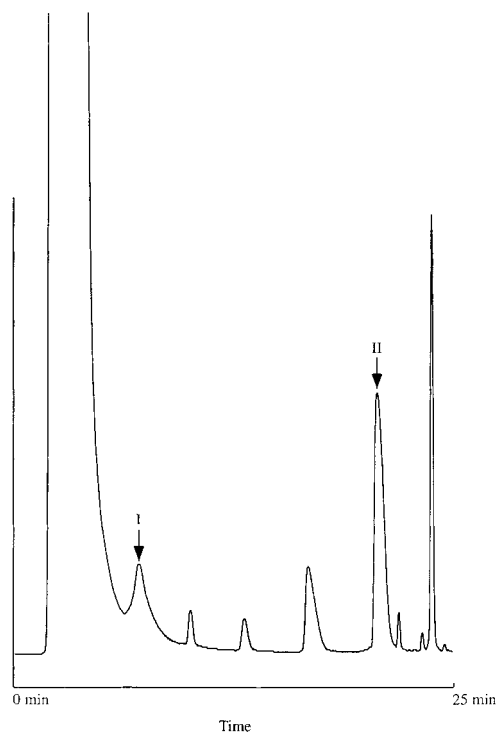


Figure 4 HPLC elution profile of products from the oxidative acylation of 25 mM phenylphosphate using 200 mM thioacetic acid in the presence of 400 mM ferricyanide after 12 h. The reaction was carried out in MES buffer (0.5 M) at pH 7.4. The products were analysed on a C18 reverse-phase HPLC column with a linear gradient of 0–60% acetonitrile (0.1% TFA) at flow rate of 1.0 ml min⁻¹ for 25 min. The eluate was monitored at 250 nm. Peak I, phenylphosphate; peak II, acetylphenylphosphate. The yield of phenylphosphate was determined by first collecting the material corresponding to each of the peaks in the elution profile and then determining the fraction of the total absorption at 250 nm present in peak II.

recent reports^{10,11} The reaction that we describe, although not regiospecific, may still find application in peptide synthesis or peptide ligation in aqueous solution but is likely to be of greatest interest in discussions of the origins of life.

If nitriles were formed abundantly in a reducing atmosphere on the primitive Earth as has often been claimed, they could have provided the starting materials for the production of thioacids. The addition of H₂S to nitriles in organic solvents is a well recognized procedure for the synthesis of thioamides¹². The reaction would presumably proceed in aqueous solution, although the yields might be lower. Thioamides are hydrolysed in aqueous solution to thioacids in good yield^{13,14}. The thioacids are stable in aqueous solution at neutral pHs¹⁵, and so could have accumulated on the primitive Earth. A simple oxidizing agent could also have formed readily on the primitive Earth, for example Fe³⁺ by photooxidation of Fe²⁺ (ref. 16). Oxidative acylation eliminates the need for an efficient prebiotic dehydrating agent, thus avoiding many of the problems associated with other proposed methods of acylation.

The assumption that the atmosphere of the Earth was once strongly reducing has been criticized¹⁷. One alternative source of reduced carbon that has been proposed is reduction of carbon monoxide or carbon dioxide on the surface of transition-metal sulphides¹⁸. Recent experimental evidence suggests that thioacetic acid is the intermediate in the efficient synthesis of acetic acid (and its activated derivatives) from carbon monoxide on FeS/NiS (ref. 18). It is not clear whether oxidative acylation with thioacetic acid formed in this way is feasible in a prebiotic context, but it is a possibility that should be investigated. □

Received 5 February; accepted 7 July 1997.

- Hulshof, J. & Ponnampemuma, C. Prebiotic condensation reactions in an aqueous medium: a review of condensing agents. *Orig. Life* **7**, 197–224 (1976).
- Ferris, J. P. in *Cold Spring Harbor Symposia on Quantitative Biology* 29–35 (Cold Spring Harbor Lab. Press, Cold Spring Harbor, NY, 1987).
- Hawker, J. R. J. & Oro, J. Cyanamide mediated syntheses of peptides containing histidine and hydrophobic amino acids. *J. Mol. Evol.* **17**, 285–294 (1981).
- Weber, A. L. & Orgel, L. E. The formation of peptides from glycine thioesters. *J. Mol. Evol.* **13**, 193–202 (1979).
- Weber, A. L. Formation of pyrophosphate, tripolyphosphate, and phosphorylimidazole with the thioester, *n*,*s*-diacetyl-cysteamine, as the condensing agent. *J. Mol. Evol.* **18**, 24–29 (1981).
- de Duve, C. *Blueprint for A Cell: The Nature and Origin of Life* (Patterson Publishers, Carolina Biological Supply Co., Burlington, NC, 1991).
- Keller, M., Blöchl, E., Wächtershäuser, G. & Stetter, K. O. Formation of amide bonds without a condensation agent and implications for origin of life. *Nature* **368**, 836–838 (1994).
- Wieland, T. in *The Roots of Modern Biochemistry: Fritz Lipmann's Squiggle and its Consequences* (eds Kleinkauf, H., von Dohren, H. & Lothar, J.) 213–221 (de Gruyter, Berlin, 1988).
- Sheehan, J. C. & Johnson, D. A. The synthesis and reactions of *N*-acyl thiol amino acids. *J. Am. Chem. Soc.* **74**, 4726–4727 (1952).
- Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. H. Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994).
- Liu, C., Rao, C. & Tam, J. P. Acyl disulfide-mediated intramolecular acylation for orthogonal coupling between unprotected peptide segments. Mechanism and application. *Tetrahed. Lett.* **37**, 933–936 (1996).
- Smith, P. A. S. *The Chemistry of Open-chain Organic Nitrogen Compounds* 165–166 (Benjamin, New York, 1965).
- Rosenthal, D. & Taylor, T. I. A study of the mechanism and kinetics of the thioacetamide hydrolysis reaction. *J. Am. Chem. Soc.* **79**, 2684–2690 (1957).
- Butler, E. A., Peters, D. G. & Swift, E. H. Hydrolysis reactions of thioacetamide in aqueous solutions. *Anal. Chem.* **30**, 1379–1383 (1958).
- Cefola, M., Peter, S. S., Gentile, P. S. & Celiano, R. A. V. Rate of hydrolysis of thioacetic acid in basic solution. *Talanta* **9**, 537–542 (1962).
- Brateman, P. S., Cairns-Smith, A. G. & Sloper, R. W. Photo-oxidation of hydrated Fe²⁺—significance for banded iron formations. *Nature* **303**, 163–164 (1983).
- Whittet, D. C. B. Is extraterrestrial organic matter relevant to the origin of life on earth? *Orig. Life Evol. Biosphere* **27**, 249–262 (1997).
- Huber, C. & Wächtershäuser, G. Activated acetic acid by carbon fixation on (Fe, Ni)S under primordial conditions. *Science* **276**, 245–247 (1997).

Acknowledgements. We thank A. R. Hill for technical assistance and S. Bailey for manuscript preparation. This work was supported by NASA.

Correspondence and requests for materials should be addressed to L.E.O.

RNA-catalysed carbon-carbon bond formation

Theodore M. Tarasow, Sandra L. Tarasow & Bruce E. Eaton

NeXstar Pharmaceuticals, Inc., 2860 Wilderness Place, Boulder, Colorado 80301, USA

The 'RNA world' hypothesis^{1–3}, which assumes that the chemical processes that led to the appearance of life were carried out by RNA molecules, has stimulated interest in catalytic reactions involving oligonucleotides such as catalytic RNA (ribozymes)⁴. Naturally occurring ribozymes have, for example, been shown to efficiently catalyse the formation and cleavage of nucleic-acid phosphodiester bonds^{4–8}, and this narrow range of RNA-catalysed reactions has been subsequently expanded by *in vitro* selection methods to include ester⁹ and amide^{10,24} bond formation S_N2 reactions and porphyrin metallations^{11,12}. Carbon-carbon bond formation and the creation of asymmetric centres are both of great importance biochemically, but have not yet been accomplished by RNA catalysis. A widely used reaction that creates two new carbon-carbon bonds and up to four stereo-centres is the Diels-Alder cycloaddition, which occurs between a 1,3-butadiene and an alkene. Here we report the successful application of *in vitro* selection to isolate pyridine-modified RNA molecules that catalyse a Diels-Alder cycloaddition. We find that the RNA molecules accelerate the reaction rate by a factor of up to 800 relative to the uncatalysed reaction.

The Diels-Alder reaction investigated in this study is shown in Fig. 1, which also shows the modified reactants. The *in vitro* selection (SELEX) for RNA molecules with Diels-Alder activity

(DAase activity) was carried out with a library of $\sim 10^{14}$ unique sequences. The RNA molecules were constructed of a contiguous 100-nucleotide randomized region, flanked by constant sequence segments to allow for amplification and other enzymatic processes. The RNA was modified by substituting 5-pyridylmethylcarboxamid-UTP¹³ (compound **2** in Fig. 1) for UTP in the transcription

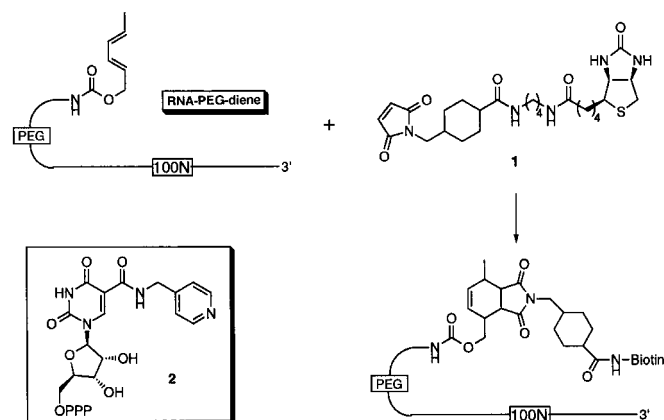


Figure 1 The Diels-Alder reaction between the acyclic diene conjugated to the RNA through a long PEG linker and the maleimide dienophile **1** (BMCC). The RNA library was prepared with a 100-nucleotide random region (100N). Biotin represents the portion of BMCC not shown in the cycloadduct. Also shown is the structure of the pyridyl methyl *n* modified UTP (**2**) substituted for native UTP during transcription; OPPP represents the triphosphate moiety.

reaction. Pyridyl-modified uridine **2** was chosen to augment the hydrogen-bonding groups of native RNA, to furnish additional hydrophobic and dipolar interactions, and to provide metal coordination sites unlike any contained in unmodified RNA. Previous attempts at isolating RNA DAases were unsuccessful using unmodified oligonucleotide libraries¹⁴. To allow for selection based on the ability to perform the desired chemistry, RNAs were coupled to an acyclic diene through a long polyethylene glycol (PEG; MW average relative molecular mass (M_r) 2,000) linker. The flexible PEG linker was used to provide the diene with the opportunity to access the RNA surface and mimic a diene substrate free in solution.

Rounds of *in vitro* selection for DAase activity were conducted by incubating the RNA-diene construct with the maleimide dienophile (compound **1** in Fig. 1; BMCC) in the presence of transition metals that could form pyridyl-Lewis acid complexes and/or coordinate to the RNA to enhance tertiary structure. The maleimide was attached to biotin so that RNA DAases could be partitioned away from unreacted RNA using streptavidin binding and denaturing polyacrylamide electrophoresis.

Following 12 rounds of *in vitro* selection, the amount of Diels-Alder reaction observed in the presence of the RNA increased from 2.4% in 8 h to 3.3% in 3 min. Subsequently, the library was cloned and bidirectionally sequenced. From the 46 sequences obtained, eight non-clonally derived families were identified. One family comprised 59% of the library population while the other seven unique sequences ranged in representation from one to six of the 46 isolates. The random region sequences for representative isolates are shown in Fig. 2a. Computer-assisted local sequence alignment of representatives from each clonal family identified one consensus sequence, UUCUAACGCG, in five of the eight non-clonally derived

a

Family 1 (27 members, 59%)

DA-1 GUGAGUGUCUUGUAGAGCGCGUGAAGAGCUUACGACUCUUCGUGUCCGUUUUCUAACGCGUGUCCAUAGAACCCGUCUCCUAGUGGGGGGGCGCCUG

DA-14 CUGAGUGUCUUGUAGAGCGCGUGAAGAGCUUACGACUCUUCGUGUCCGUUUUCUAACGCGUGCCCAUGAACCCUCCUAGUUGGGGGGGCGCCUG

DA-22 UUGAGUGUCUUGUAGAGCGCGUGAAGAGCUUACGACUCUUCGUGUCCGUUUUCUAACGCGUGCCCAUGAACCCUCCUAGUUGGGGGGGCGCCUG

DA-47 UUGAGUGUCUUGUAGAGCGCGUGAAGAGCUUACGACUCUUCGUGUCCGUUUUCUAACGCGUGCCCAUGAACCCUCCUAGUUGGGGGGGCGCCUG

Family 2 (6 members, 13%)

DA-2 GGCUUUCCAGUAUGGAUACCCUUGGGGCUUUGUAGAGAGCGUAGUUGUUUCUAACGCGUAGGCCUGGUGAGGUACGUGUAGAUGCGGUUAUCGAGGG

Family 3 (4 members, 9%)

DA-37 AGGUAGAAUGGUGUCCAGUGUGUAUAACAAAGGUGUCGGGUGACCGGUAUCGUCUGUCCUUGGCGUGGCAUUGUGGGUGUGUCGACUCGCAAUCCG

Family 4 (4 members, 9%)

DA-17 GGGUUGCGUGACUUGUAGAAGCGUAGCUUUUCUAACGCGGCGGUGUACCGUUGGAGCUCUCGAGUUGGGUCCUUGCCUUAUGGUGAAGCGG

Family 5 (2 members, 4%)

DA-24 ACGCGCAUCUAAGUGGUCUCCGUCUCGUCACGAGUAUCGUGUCCUUGGCUUUGAGUCUCUGUGACUAGGACUAGGGACGUCGGGUCUGGGGA

DA-7 CGCAUCGGGCGUGGACGUGUCUUGUGGAUUGUUCUGGCGAAUGGCCGGGAAUUUCUAACGCGCGCACCUCUGUUGGGGAGCAUUGGUGCGAUGAG

DA-10 UCCACACAAAACGGUGGCGAUAGACCUUGUAGACAAUUGCCGGCGAAAAGAGAGUCGGUUGUUGUUUCUAACGCGAGGCGGGGUGCUAAUAGGGCC

DA-15 GGUGUCGUGUACGUAUCGUCUGUGCCAGGAGUCCUUCUGUGUACGGAAGUCGAUUGCUAGCUCGCGCGCGGAAUUGCGGAUUCGGUGUGGCA

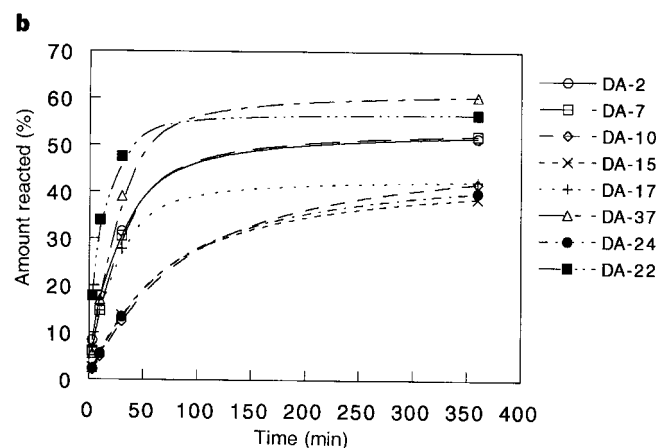


Figure 2 a, Random region ($\sim 100N$) sequences of 11 isolates obtained from the DAase *in vitro* selection. Isolates are identified by the number to the left of the sequences. Members of clonal families are labelled (in parentheses) with the total number of family members and the family population as a percentage of the total number of sequences. The computer-identified consensus sequence is in bold and underlined for each of the isolates from families 1, 2 and 4 and the two orphans. **b**, Percentage of individual RNA isolates reacted as a function of time. Isolates are represented by the symbols listed to the right and were incubated with 500 μ M BMCC for the times indicated.

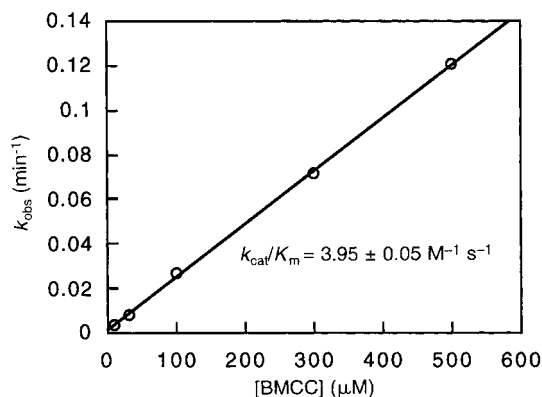


Figure 3 The observed rate constant (k_{obs}) increases as a function of BMCC concentration for isolate 22. A linear fit of the data yields the line shown ($y = 0.001545x + 0.0002373x$; $R = 0.9998$) with the ratio k_{cat}/K_m determined from the slope.

sequences analysed (Fig. 2a). Beyond this limited amount of homology, there are no obvious sequence or structural similarities between the eight non-clonally derived families.

Eight of the isolates shown in Fig. 2a were analysed for their DAase activity. Isolates were kinetically evaluated under solution-saturating concentrations of BMCC (500 μM). All other conditions were identical to those used during the selection. The results are summarized in Fig. 2b and indicate that each of the isolates tested was capable of facilitating the Diels–Alder reaction, establishing that there were at least eight unique sequences present in the original library that could promote this cycloaddition reaction.

Isolate 22 (DA-22) was chosen for more detailed characterization. Control experiments demonstrated that reaction product formation required both substrates (diene and dienophile). DAase activity was completely dependent on the presence of the pyridyl-modified uridine, as RNA transcribed using the native nucleotide triphosphate was inactive. This is not surprising as the unmodified RNA may have folded into considerably different three-dimensional structures. In addition, the absence of the pyridine groups could have eliminated unique metal-binding sites that effect Lewis acids catalysis.

The insolubility of the dienophile substrate, BMCC, prevented complete Michaelis–Menten kinetic analysis of isolate 22. Nevertheless, kinetic analysis was performed on isolate 22 up to a maximum concentration of 500 μM BMCC (Fig. 3). The data are linear over the range of dienophile concentrations used with a slope equal to k_{cat}/K_m ($3.95 \pm 0.05 \text{ M}^{-1} \text{ s}^{-1}$), where k_{cat} is the catalytic rate constant and K_m is the apparent dissociation constant of the RNA–substrate complex. Comparing the apparent second-order rate constant to the uncatalysed rate constant k_{uncat} of $5.42 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$ indicates that isolate 22 achieved an 800-fold rate acceleration of the Diels–Alder cycloaddition. The uncatalysed rate was measured using an identical construct comprised of random pyridine-modified RNA.

Lineweaver–Burke analysis of the data from 30 to 500 μM BMCC gave estimates of k_{cat} ($0.011 \pm 0.002 \text{ s}^{-1}$) and K_m ($2.3 \pm 0.5 \text{ mM}$). Comparison of k_{cat} to the second-order rate constant for the uncatalysed reaction, k_{uncat} , indicates that isolate 22 has an effective

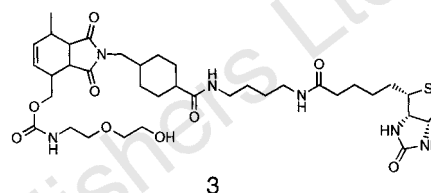
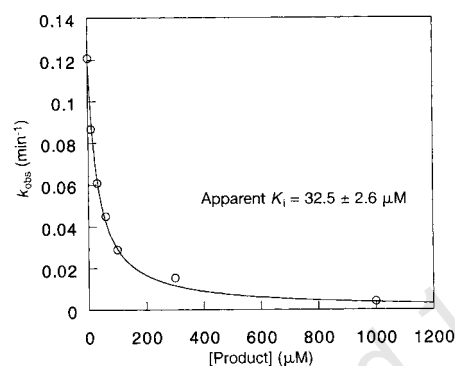


Figure 4 Inhibition of the DAase activity of isolate 22 by the free cycloaddition product **3** with an apparent K_i shown. Observed rate constants were measured at increasing concentrations of **3** in the presence of 500 μM BMCC (**1**) and fitted to the equation described in Methods.

molarity of the order of 2 M (ref. 15). Despite a K_m 10-fold higher than the concentration of BMCC (100 μM) used in the *in vitro* selection, the sequences from Family 1 (see Fig. 2a) seem to have been selected for their ability to increase the reaction rate of bound reactants (k_{cat}) and in effect compensated for a high K_m . These results clearly indicate that saturating concentrations of substrate are not necessary for the successful selection of RNA catalysts.

Predictably, the free product of the Diels–Alder reaction acts as a reasonably good inhibitor for the RNA-catalysed reaction (Fig. 4). Again, the insolubility of BMCC precluded determination of an absolute inhibition constant (K_i), for the product but data collected at 500 μM BMCC and varying amounts of the product (compound **3** in Fig. 4) were used to determine an apparent K_i of 32.5 μM (ref. 16). Inhibition by the product is consistent with the RNA mediated Diels–Alder reaction occurring in a binding pocket which resembles the product or at least can readily undergo conformational changes to bind the product of the catalysed reaction.

The metal dependence of isolate 22 indicates an absolute dependence on Cu^{2+} . No DAase activity was observed in the absence of Cu^{2+} even in the presence of the other divalent metals used in the *in vitro* selection. Magnesium, calcium and copper restored DAase activity to ~71% of that achieved by the complete metal mixture. The additional 29% incremental improvement appears to be a nonspecific metal–RNA interaction that can be accomplished using any of the divalent metals found in the reaction buffer. Titration experiments indicated maximum activity at 10–20 μM Cu^{2+} , a level similar to that used in the *in vitro* selection (10 μM). The absolute and specific metal dependence on copper for DAase activity suggests that copper Lewis acid sites are formed as other metal ions present in the reaction buffer could have adopted similar coordination geometries^{17,18} resulting in similar RNA structures. The selection of copper in this capacity is consistent with known Lewis-acid-catalysed Diels–Alder reactions in water^{19,20}. Indeed, although divalent metal ions have been shown to play an important role in the activity of other oligonucleotide catalysts, such exclusive dependence on Cu^{2+} has not been observed.

We used mass spectrometry to identify the DAase reaction product, in order to firmly establish RNA DAase activity^{21–23}.

Following reaction of isolate 22 with BMCC, the RNA was isolated by gel filtration and digested with ribonuclease I. The sample was then purified by high-performance liquid chromatography (HPLC) and subjected to electrospray-ionization tandem quadrupole mass spectrometry. From the resulting data, the ion corresponding to the Diels–Alder product was identified (measured M_r 630.6; calculated M_r 630.8). Moreover, many additional fragment ions were observed which further substantiate formation of the Diels–Alder product. No ions were observed for BMCC reacting with any functional groups on the oligonucleotide, consistent with only the formation of the Diels–Alder cycloadduct.

The methodology used to create these RNA DAases provides a straightforward approach to generating novel catalysts on demand that do not require templating of either substrate. The scope of RNA-catalysed reactions now includes carbon–carbon bond-forming reactions. Although no examples of RNA-catalysed carbon–carbon bond formation were previously known, there appear to be many solutions to $[4 + 2]$ cycloaddition catalysis, as eight unique sequences were found that enhance the rate of the Diels–Alder reaction. These results suggest that similar strategies could be used to identify RNA molecules that catalyse other Diels–Alder reactions, including those with typically unreactive substrates, inverse electron demand Diels–Alder cycloadditions, and hetero Diels–Alder reactions. RNA catalysis of other types of cycloaddition reactions such as dipolar cycloadditions, particularly those benefiting from Lewis acid catalysis, are also possibilities. The ability to expand greatly the functional diversity of RNA through modified bases, to augment accessible chemistries through the use of transition metals, and, perhaps in the future, to include co-factor-assisted transformations, has significant implications for the range of reactions amenable to RNA catalysis. □

Methods

Incubation conditions. The RNA–PEG–diene construct was prepared by ligating on a PEG–diene modified DNA 10-mer to the 5' end of the RNA using T4 DNA ligase. All RNA incubations were conducted under the following conditions except as noted: 50 mM HEPES, pH 7.0, 500 nM pyridyl methyl-modified RNA, 200 mM NaCl, 200 mM KCl, 1 mM CaCl_2 , 1 mM MgCl_2 , 10 μM each aluminium lactate, $\text{Ga}_2(\text{SO}_4)_3$, MnCl_2 , FeCl_2 , CoCl_2 , NiCl_2 , CuCl_2 and ZnCl_2 , 10% ethanol and 2% dimethyl sulphoxide. The concentration of dienophile 1 (BMCC) varied in the isolate characterization experiments, but was held constant at 100 μM throughout the SELEX. Incubations were terminated by the addition of β -mercaptoethanol to a final concentration of 5 mM and/or passing the solution over two successive Nap columns (Pharmacia) to remove excess BMCC.

Reaction assay and partitioning. The extent of reaction and partitioning of reacted and unreacted RNA molecules was accomplished using a streptavidin (SA) dependent gel shift. The shifted and unshifted bands were visualized by autoradiography and phosphor-imaging, the latter being used for quantification. For partitioning, shifted bands were excised, the RNA–SA complex extracted, desalted and subjected to reverse transcription and PCR amplification according to standard procedures.

Kinetic analyses. All data were obtained at 500 nM RNA and the indicated amounts of BMCC. k_{obs} values were determined by fitting the fraction of unreacted RNA to the equation for first-order kinetics. The uncatalysed second-order rate of Diels–Alder reaction was measured using random pyridine-modified RNA ($k_{\text{uncat}} = 5.42 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$).

Product inhibition. Apparent K_i values for the free cycloaddition product 3 were determined at 500 μM BMCC by fitting the observed first-order rate constants to the following equation for inhibition: $k_{\text{obs}} = (k_{\text{obs0}}/2)(\alpha E - I - K_i + ((K_i + \alpha E - I)^2 + 4K_i)^{1/2})$ where k_{obs} is the measured rate constant in the presence of 3, k_{obs0} is the observed rate constant in the absence of 3, αE represents the fractional (α) concentration of functional active sites (E), I is the concentration of 3, and K_i is the apparent inhibition constant.

Received 23 April; accepted 22 July 1997.

- Joyce, G. F. Ribozymes: Building the RNA world. *Curr. Biol.* **6**, 965–967 (1996).
- Joyce, G. F. The rise and fall of the RNA world. *New Biologist* **3**, 399–407 (1991).

- Joyce, G. F. Some biochemical thoughts on the RNA world. *Chem. Biol.* **3**, 405–407 (1996).
- Cech, T. R. The chemistry of self-splicing RNA and RNA enzymes. *Science* **236**, 1532–1539 (1987).
- Long, D. M. & Uhlenbeck, O. C. Self-cleaving catalytic RNA. *FASEB J.* **7**, 25–30 (1993).
- Bartel, D. P. & Szostak, J. W. Isolation of new ribozymes from a large pool of random sequences. *Science* **261**, 1411–1418 (1993).
- Beaudry, A. A. & Joyce, G. F. Directed evolution of an RNA enzyme. *Science* **257**, 635–641 (1992).
- Kumar, P. K. R. & Ellington, A. D. Artificial evolution and natural ribozymes. *FASEB J.* **9**, 1183–1195 (1995).
- Illangasekare, M., Sanchez, G., Nickles, T. & Yarus, M. Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **267**, 643–647 (1995).
- Lohse, P. A. & Szostak, J. W. Ribozyme-catalyzed amino-acid transfer reactions. *Nature* **381**, 442–444 (1996).
- Li, Y. & Sen, D. A catalytic DNA for porphyrin metallation. *Nature Struct. Biol.* **3**, 743–747 (1996).
- Conn, M. M., Prudent, J. R. & Schultz, P. G. Porphyrin metallation catalyzed by a small RNA molecule. *J. Am. Chem. Soc.* **118**, 7012–7013 (1996).
- Dewey, T. M., Zyzanski, C. & Eaton, B. E. The RNA world: functional diversity in a nucleoside by carboxyamidation of uridine. *Nucleosides & Nucleotides* **15**, 1611–1617 (1996).
- Morris, K. N. et al. Enrichment for RNA molecules that bind a Diels–Alder transition state analog. *Proc. Natl Acad. Sci. USA* **91**, 13028–13032 (1994).
- Jencks, W. P. in *Catalysis in Chemistry and Enzymology* 644–712 (Dover Publications, New York, 1987).
- Williams, J. W. & Morrison, J. F. The kinetics of reversible tight-binding inhibition. *Methods Enzymol.* **63**, 437–467 (1979).
- Kazakov, S. A. in *Bioorganic Chemistry: Nucleic Acids* (ed. Hecht, S. M.) 244 (Oxford Univ. Press, New York, 1996).
- Cotton, F. A. & Wilkinson, G. *Advanced Inorganic Chemistry* (Wiley, New York, 1988).
- Otto, S., Bertoncin, F. & Engberts, J. B. F. N. Lewis acid catalysis of a Diels–Alder reaction in water. *J. Am. Chem. Soc.* **118**, 7702–7707 (1996).
- Otto, S. & Engberts, J. B. F. N. Lewis acid catalysis of a Diels–Alder reaction in water. *Tetrahedr. Lett.* **36**, 2645–2648 (1995).
- Ni, J., Pomerantz, S. C., Rozenski, J., Zhang, Y. & McCloskey, J. M. Interpretation of oligonucleotide mass spectra for determination of sequence using electrospray ionization and tandem mass spectrometry. *Anal. Chem.* **68**, 1989–1999 (1996).
- Pomerantz, S. C., McCloskey, J. A., Tarasow, T. M. & Eaton, B. E. Deconvolution of combinatorial oligonucleotide libraries by electrospray ionization tandem mass spectrometry. *J. Am. Chem. Soc.* **119**, 361–3867 (1997).
- Tarasow, T., Tinnermeier, D. & Zyzanski, C. Characterization of oligodeoxyribonucleotide-polyethylene glycol conjugates by electrospray mass spectrometry. *Bioconjugate Chem.* **8**, 89–93 (1997).
- Wiegand, T. W., Janssen, R. C. & Eaton, B. E. Selection of RNA amide synthases. *Chem. Biol.* (in the press).

Acknowledgements. We thank L. Gold for support, guidance and vision; the scientific community at NeXstar, especially members of the Medicinal Chemistry group, for helpful discussions and ideas; S. Wayland for the synthesis of compound 2; and T. Wiegand and D. Nieuwlandt for inspiring dialogue and technical assistance. We also thank S. C. Pomerantz, P. F. Crain and J. A. McCloskey for ESI–MS/MS analysis.

Correspondence and requests for materials should be addressed to B.E.E. (e-mail: beaton@nexstar.com).

Abrupt mid-twentieth-century decline in Antarctic sea-ice extent from whaling records

William K. de la Mare

Australian Antarctic Division, Department of the Environment, Sport and Territories, Channel Highway, Kingston, Tasmania 7050, Australia

A decline in Antarctic sea-ice extent is a commonly predicted effect of a warming climate. Direct global estimates of the Antarctic sea-ice cover from satellite observations, only possible since the 1970s^{1–4}, have shown no clear trends. Comparisons¹ between satellite observations and ice-edge charts obtained from early ship records⁵ suggest that sea-ice extent in the 1970s was less than during the 1930s, an indication supported by limited regional observations⁶. But these observations have been regarded as inconclusive, owing to the limited spatial and temporal scope of the early records². A significant data source has, however, been overlooked. The southern limit of whaling was constrained by sea ice, and since 1931 whaling records have been collected for every whale caught⁷, giving a circumpolar coverage from spring to autumn until 1987. Here, an analysis of these catch records indicates that, averaged over October to April, the Antarctic summer sea-ice edge has moved southwards by 2.8° of latitude between the mid 1950s and early 1970s. This suggests a decline in the area covered by sea ice of some 25%. This abrupt change poses a challenge to model simulations of recent climate change, and could imply changes in Antarctic deep-water formation and in biological productivity, both important processes affecting atmospheric CO₂ concentrations.

Whaling in the Antarctic began in 1904 from land stations and in 1905 the first floating factories were introduced. The early factory ships lacked slipways for hauling whales on board, but they could process the catch by mooring to an ice floe, leading to the fortuitous discovery that whales, especially the highly prized blue whales, tended to concentrate near the ice edge⁷. The sea-ice margin is an area of enhanced biological productivity^{8,9}, and pelagic whaling concentrated near the ice edge for most of the commercial era. The introduction of floating factories with stern slipways led to the expansion of whaling around Antarctica, and by 1931 the pattern of whaling near the ice edge was well established¹⁰. The whaling season usually began in October and continued through the Antarctic summer until April. The whaling fleets spread along the ice edge at the start of the season, following it southwards as it retreated¹⁰. No pelagic whaling occurred from 1940 to 1946, but after the Second World War catching again concentrated near the ice edge. From 1957 to 1975, the scarcity of blue, fin and humpback whales led to whaling concentrating on sei whales in the region of the South Polar Front, well north of the ice edge¹¹. In 1972 whaling turned to the remaining abundant species, the relatively small minke whale. As with blue, fin and humpback whaling, minke whaling occurred near the ice edge¹², and became widespread until the end of commercial whaling after 1986/87.

The International Whaling Commission has entered about 1.5 million catch records into computer files from logbooks submitted to the Bureau of International Whaling Statistics in Norway. The data recorded for each whale include the species, date of capture and the noon position of the factory ship (to at least the nearest degree of latitude and longitude). I used a database program to extract the records for the southernmost catch for each unique longitude and date. The data from land stations at South Georgia were excluded, as were catches after the 1956/57 season for species other than minke whales. This resulted in 42,258 records. No records satisfied the selection criteria from 1960 to 1971 inclusive, when sei whaling

predominated. The mean latitudes of the southernmost catches were calculated for each combination of three independent factors; the 10° longitudinal sector ('sector'), 10-day period ('decade') and season ('year'). Three decades within each month were defined by the days 1–10, 11–20, >20. The data were censored to exclude any positions more than 3° north of the southernmost position for each factor combination. There were 38 seasons, 19 decades and 36 sectors, leading to 25,992 combinations. Not all combinations are represented in the data; 4,424 combinations have an observed mean southernmost catch latitude. The seasons were in three sections, 1931/32 to 1940/41, 1946/47 to 1958/59 and 1971/72 to 1986/87. The dates within seasons ranged from 14 October to April 20.

An analysis of variance was carried out using a generalized linear model with an identity link function and normal errors¹³, with the mean southernmost catch latitude as the dependent variable. Only main effects were estimated because there is only one observation per cell. The observations cross the levels of the three factors sufficiently for all the coefficients of the linear model to be estimatable. The analysis of variance table (Table 1) shows that each factor is statistically significant.

The mean effect due to a given factor can be plotted as the predicted value of the dependent variable, and its standard error, for each level of the given factor, with the other factors held fixed. Figure 1 shows the predicted mean latitude of the southernmost catches by year. The mean latitudes for the southernmost catches were roughly constant in the period 1931–54 with a mean of 61.5° S. In the mid 1950s the catch latitudes begin to move southward, although the pattern of the change cannot be estimated during the period of sei whaling. From 1973 onwards the latitude is again roughly stable, but 2.8° further south, with a mean of 64.3° S.

A plot of the average catch latitude by time of year showed a smooth curve corresponding to the seasonal pattern of the retreat and advance of the sea ice, with a minimum estimated to occur in the first decade in March (see Supplementary Information). Figure 2

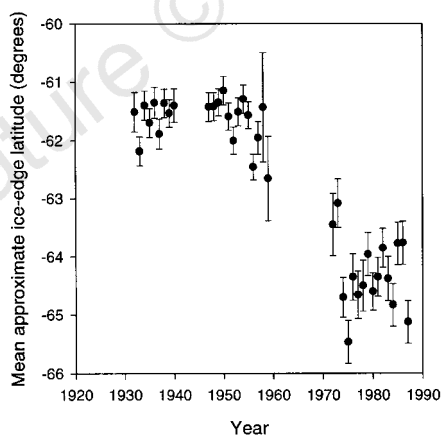


Figure 1 Mean approximate ice-edge latitude in Antarctica from 1931 to 1987. The estimates are from a linear model fitted to whale-catch records, and are standardized to first 10-day period of January and the longitudinal sector 20°–30° E. The year is defined by the first decade (ten days) in January, and so, for example, 1932 is the mid-point of the 1931/32 season. The predictions are most precise at the centre-of-mass of the data, and so the selected decade and longitude are based on the mean value of each factor weighted by the number of observations at each of its levels. The year effects are corrected for decade and longitude, and so describe a generalized effect for the latitude of catches with time. The actual pattern over time for a given sector and decade would not necessarily correspond exactly to the pattern shown because there are likely to be interactions between the factors. However, these could not be estimated with the data available. The error bars represent ± 1 standard error.

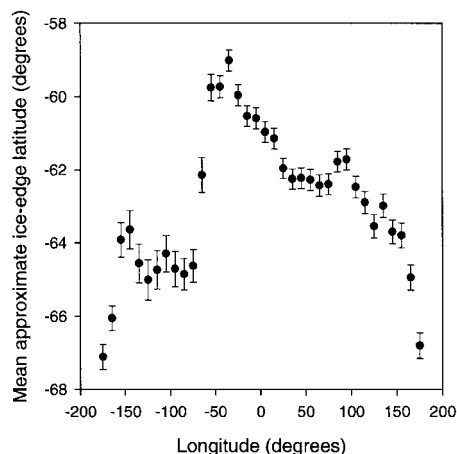


Figure 2 Mean approximate ice-edge latitude in Antarctica by 10° longitudinal sector. The estimates are from a linear model fitted to whale-catch records, and are standardized to the first decade of January and the year 1957. Longitudes to the west are plotted as negative values. The southernmost catches are around 180°W, corresponding to the middle of the Ross Sea. At 70°W the mean catch latitude moves north around the Antarctic Peninsula, and remains at a low latitude across the Scotia Sea. The latitudes move steadily southwards, showing a dip into Prydz Bay at 65°–75° E, and a small bulge at 85° and 95° E corresponding to the West Ice Shelf and Shackleton Ice Shelf, before continuing to trend southwards into the Ross Sea. The error bars represent ± 1 standard error.

shows the longitudinal pattern in the latitudes of the southernmost catches, which reflects the outline of the sea ice around Antarctica. These results show that the analysis captures the major features of the Antarctic sea ice, lending confidence that the pattern in Fig. 1 is a real reflection of trends in Antarctic sea-ice coverage.

Interpretation of Fig. 1 can be improved by analysing the relationship between the noon positions of the factory ships and the ice edge, for which two additional data sets are available: charts of the sea-ice edge⁵ and data compiled³ from Joint Ice Center (JIC) charts, based largely on satellite data. Figure 3a shows the strong correlation between the mean southernmost catch latitudes with the ice-edge data⁵ for the 'years' 1932–39. This ice edge is described as the "close-pack", and is therefore likely to be sea-ice coverage of 80%

Table 1 Analysis of variance for the southernmost catch positions of whales in the Antarctic

Factor	Sum of squares	Degrees of freedom	Mean square	F value	P
Sector	19,688.18	35	562.52	112.97	<0.000001
Decade	22,224.24	18	1,234.68	247.95	<0.000001
Year	4,581.13	37	123.81	24.86	<0.000001
Residual	21,576.43	4,333	4.98		

The southernmost catch position is the dependent variable; $R^2 = 0.683$. The factors are defined in the text.

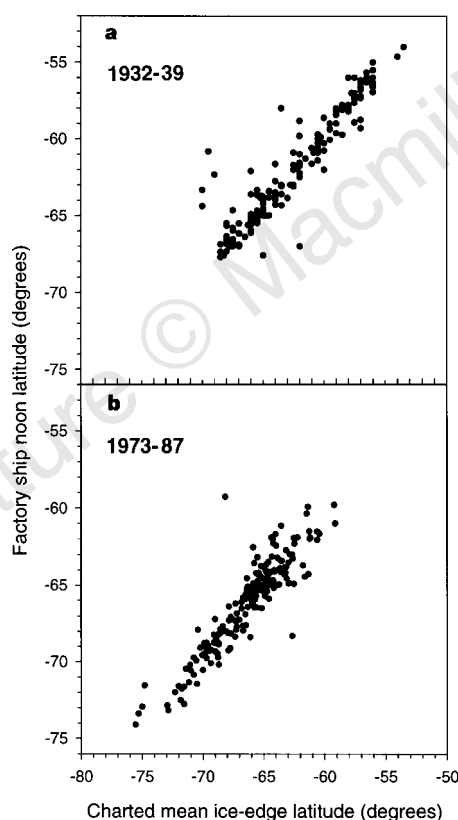


Figure 3 The relationship between the latitudes of the southernmost whale catches, in terms of the factory-ship noon positions, and information on the ice-edge given in charts published in reports⁵ of the Discovery Committee (a) and from data³ derived from charts published by the Joint Ice Center (JIC) (b). The Discovery data cover the 'years' 1932–39, and the JIC data cover 1973–87. There are 178 observations where the Discovery and southernmost-catch data can be compared. A linear regression of the catches on the Discovery data gives $R^2 = 0.88$, with a slope of 0.845 (standard error, 0.024). There are 196 observations where the JIC and southernmost-catch data sets can be compared. The regression of catch positions on the JIC data gives $R^2 = 0.832$, with a slope of 0.875 (standard error, 0.028).

or more. The mean difference between this ice edge and the southernmost catch latitudes is 0.58° (35 nautical miles) with standard error 0.11° (6.5 nautical miles), which is statistically significant ($t = 5.38$, $P < 0.001$). The analysis includes data collected by whaling fleets, which ensures a high degree of correlation, but also constitutes direct evidence of their close proximity to the ice edge. Figure 3b shows the strong correlation between the catch latitudes and the JIC ice-edge data for the period 1973–87. The mean difference between the latitude of the southernmost catches and the JIC ice edge is 0.14° (8.3 nautical miles) to the north (standard error, 0.09°). The whaling latitudes are closer because the JIC ice edge used applies to 15% coverage.

One possible explanation for the change in whaling latitudes is that it is entirely related to the differences in species composition between the blue–humpback–fin era and the later minke era. The use of only the southernmost catches, their high correlation with direct ice-edge observations, the direct descriptions of the relationship between whaling and the ice edge^{7,10,12} and knowledge of the distributions of whales obtained during sightings surveys¹⁴ all show that this explanation can be conclusively rejected. Statistical confounding can also be rejected; there is no systematic shift of any consequence in longitudinal or decadal coverage between the blue–fin–humpback era and the minke era (see Supplementary Information).

The difference between the mean distances from the ice edge to factory-ship positions for the blue–humpback–fin era and the later minke era is almost certainly due largely to the differences in the definitions of the ice edge used in the charts and the compiled³ JIC data. There was no change in whaling technology between the 1950s and 1980s that would allow whaling in heavier ice concentrations in the more recent period. Whale catching requires water that is substantially clear of ice. The positions of factory ships in the minke era are about 5–20 nautical miles north of the 'hard' ice-edge¹². Therefore, the relationship between the southernmost whale catches and the zone where ice becomes too thick for catching is substantially the same in the minke period as it was in the blue–humpback–fin period. If there is a difference, it is certainly less than 0.5° ; too small to account for the changes shown in Fig. 1. Therefore, Fig. 1 indicates real trends in Antarctic sea-ice extent over the period. The mean area of sea ice from 1973 to 1981 was² $17.4 \times 10^6 \text{ km}^2$; 2.8° of latitude at 63°S is an area of $5.65 \times 10^6 \text{ km}^2$, which corresponds to a decline in sea-ice coverage of about 25%.

Figure 1 shows that a substantial decline in sea-ice extent has occurred. Although this has been suggested before^{1,4,6}, the whale-catch records bring much more data to bear, and with circumpolar coverage. They also cover the period 1945–59, during which (apart from ref. 15) few direct observations of the ice edge were reported. The decline occurred relatively quickly, beginning in the mid 1950s, and was largely complete by 1973. Frustratingly, the change pre-dates reliable satellite observations. Based on satellite data, the International Panel on Climate Change concluded that Antarctic sea-ice coverage since 1973 had remained close to average¹⁶. These analyses also show that sea-ice extent has been roughly stable since 1973. However, the abrupt change from the 1950s to the 1970s shows that Antarctic sea-ice extent was not stable before 1973.

A system in which the sea-ice extent seems to be roughly stable and then changes abruptly to again appear roughly stable provides an interesting challenge for coupled atmosphere–ocean general circulation models; analysis of such a system may give insights into fundamental climate processes. The importance of the marginal sea-ice zone in primary production suggests that a decline in Antarctic marine productivity may have already occurred. □

Received 19 November 1996; accepted 9 July 1997.

- Kukla, G. & Gavin, J. Summer ice and carbon dioxide. *Science* **214**, 497–503 (1981).
- Zwally, H. J., Parkinson, C. L. & Comiso, J. C. Variability of Antarctic sea-ice and changes in carbon dioxide. *Science* **220**, 1005–1012 (1983).

3. Jacka, T. H. Antarctic and Southern Ocean sea-ice and climate trends. *Ann. Glaciol.* **14**, 127–130 (1990).
4. Jacka, T. H. & Budd, W. F. in *International Conference on the Role of the Polar Regions in Global Change* (eds Weller, G., Wilson, C. L. & Severin, B. A. B.) 63–70 (Univ. Alaska, Fairbanks, 1991).
5. Mackintosh, N. A. & Herdman, H. F. P. Distribution of the pack-ice in the Southern Ocean. *Discovery Rep.* **19**, 285–296 (1940).
6. Murphy, E. J., Clarke, A., Symon, C. & Priddle, J. Temporal variation in Antarctic Sea-ice: analysis of a long term fast ice record from the South Orkney Islands. *Deep-Sea Res.* **1** **42**, 1045–1062 (1995).
7. Tonneson, J. N. & Johnsen, A. O. *History of Modern Whaling* (Hurst, London, 1984).
8. Hart, T. J. On the phytoplankton of the southwest Atlantic and the Bellingshausen sea. *Discovery Rep.* **8**, 1–208 (1934).
9. Smith, W. O. J. & Nelson, D. M. Phytoplankton bloom produced by a receding ice-edge in the Ross Sea: spatial coherence with the density field. *Science* **227**, 163–167 (1985).
10. Hjort, J., Lie, J. & Ruud, J. T. Norwegian pelagic whaling in the Antarctic III. *Hvalradets Skrifter* **8**, 4–36 (1933).
11. Horwood, J. *The Sei Whale: Population Biology, Ecology and Management* (Croom Helm, London, 1987).
12. Shimadzu, Y. & Katabami, Y. A note on the information on the pack ice edge obtained by Japanese catcher boats in the Antarctic. *Rep. Int. Whal. Commis.* **34**, 361–363 (1984).
13. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S-Plus* (Springer, New York, 1994).
14. Kasamatsu, F., Joyce, G. G., Ensor, P. & Mermoz, J. Current occurrence of baleen whales in Antarctic waters. *Rep. Int. Whal. Commis.* **46**, 293–304 (1996).
15. Heap, J. *Sea Ice Distribution in the Antarctic between 7° W and 92° W* (Admiralty Hydrographic Off., London, 1963).
16. Nicholls, N. et al. in *Climate Change 1995: The Science of Climate Change* (eds Houghton, J. T. et al.) 137–192 (Cambridge Univ. Press, 1996).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial office of Nature.

Acknowledgements. I thank C. Allison and staff at the International Whaling Commission for compiling these data. I also thank I. Allison, J. Heap, H. Marchant and P. Quilty for comments.

Correspondence and requests for materials should be addressed to the author (e-mail: bill_de@antdiv.gov.au). Southernmost catch records and direct sea-ice data are available on <http://www.antdiv.gov.au/aad/sci/human/aad/c/dissemination/dissemination.html>.

Measurements of electric anisotropy due to solidification texturing and the implications for the Earth's inner core

Michael I. Bergman

Physics Department, Simon's Rock College, Great Barrington, Massachusetts, 01230, USA; and Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA

Seismic body-wave and normal-mode data^{1–4} suggest that the Earth's solid inner core is elastically anisotropic, with the fast direction nearly parallel to the rotation axis. Compressional body-wave data also suggest that the anisotropy increases with turning depth, with a maximum anisotropy of 3–4% (refs 5–7). Here I propose that the inner core's elastic anisotropy and the depth dependence of the anisotropy may be due to solidification texturing that results from the dendritic growth of iron crystals. I demonstrate through laboratory measurements that directionally solidified metallic alloys can indeed exhibit a significant elastic anisotropy due to solidification texturing. The anisotropy is due to the dendrites growing along a particular crystallographic axis^{8,9}, which tends to be aligned along the direction of heat flow. Directional cooling in the Earth's core must therefore be predominantly in the cylindrically radial direction in order for solidification texturing to cause the observed anisotropy; such directional heat flow may be consistent with the pattern of convection in the outer core¹⁰. It is possible that columnar crystals composed of dendrites and elongated in the cylindrically radial direction could also explain observations of inner-core attenuation anisotropy^{11,12}.

The elastic anisotropy of the individual crystals composing the inner core may be responsible for the observed seismic anisotropy if there is a mechanism to align the crystals and cause a preferred orientation, or texturing, of the crystals. Possible mechanisms include deformation texturing due to thermally driven solid-state

convection¹³, and texturing due to solidification in a magnetic field of crystals with an anisotropic paramagnetic susceptibility¹⁴. It has also been suggested that the entire inner core might be a single crystal, obviating the need for an alignment mechanism, and yielding a maximum anisotropy¹⁵. Another possibility is that the pattern of convection in the fluid outer core could lead to latitudinal variations in the inner-core growth rate. Isostatic adjustment via solid-state flow might then relax the inner–outer core boundary towards the equipotential, allowing the development of a deformation or recrystallization texture^{16,17}. Here I discuss another mechanism, directional solidification and the resulting solidification texturing.

When an alloy melt solidifies directionally, a flat interface can become morphologically unstable because of constitutional supercooling⁸. With increasing instability, solvent-rich dendrites form, which make up the mushy zone between liquid and solid. A small temperature gradient and a large solidification front speed favour a thick mushy zone¹⁸. As primary dendrites grow along a particular crystallographic axis^{8,9}, and also tend to grow along the direction of heat flow, a particular crystallographic axis tends to lie in the direction of heat flow, even in a polycrystalline metal^{19,20}. This is solidification texturing. For face-centred cubic (f.c.c.) and body-centred cubic (b.c.c.) metals the axis of primary dendritic growth is $\langle 100 \rangle$, for hexagonal close-packed (h.c.p.) metals the axis is $\langle 210 \rangle$, and for tetragonal metals $\langle 110 \rangle$.

It is thought that the Earth's core is an iron-rich alloy of iron (as well as some nickel) and a lighter constituent, perhaps metallic iron oxide or iron sulphide^{16,21}. Numerical estimates suggest that the inner core may grow dendritically, and that the mushy state may extend deep into the inner core²². There is also meteoritic evidence from presumed planetoid cores that hints at dendritic structures with primary arm spacing of the order of tens of metres (ref. 23), a large length scale that results from the small cooling rate and front speed²⁴. Numerical estimates also suggest that the fluid in the mushy zone beneath the inner–outer core boundary is convecting vigorously²⁵, which is predicted to result in a very high solid fraction except in the very upper reaches of the inner core²⁶. A rapid change in the solid fraction is consistent with seismic reflections off the inner–outer core boundary²⁷, while small fluid inclusions could be responsible for the high overall attenuation of seismic waves in the inner core²⁸.

To examine whether solidification texturing could yield a measurable elastic anisotropy, I studied dendritic growth in a tin-rich alloy. Although tin has a centred tetragonal structure, whereas iron under inner-core pressure and temperature is thought to have either an h.c.p. or an f.c.c. lattice^{15,29}, the solidification-texturing mechanism is essentially the same in all dendritic metallic alloys. I solidified a 97% Sn–3% Pb, 400-g, 35-mm-diameter cylindrical ingot directionally by cooling from beneath (Fig. 1). By alloying the tin with a few per cent lead (f.c.c.) constitutional supercooling and dendritic growth of tin-rich dendrites is promoted¹⁸. Because there is only a low percentage of solute in the original melt, the solidified ingot has only a small percentage of solute-enriched interdendritic material. In the Earth's inner core the percentage of solute-enriched interdendritic material is also small, even though the solute percentage of the core as a whole may be larger, because of fractionation and convection of light solute into the outer core^{22,25}. Although the pressure, temperature, composition, and length and timescales of the experiments differ widely from those in the inner core, I have modelled a high-solid-fraction dendritic zone because such a zone has been predicted to exist even under those very different conditions in the inner core²².

Tin dendrites grow along the $\langle 110 \rangle$ axis, but in the plane perpendicular to dendritic growth a given crystal can be orientated arbitrarily. Thus, even if a polycrystalline tin sample exhibits perfect alignment of the $\langle 110 \rangle$ axes along the direction of heat flow, any orientation in the plane perpendicular to the direction of heat flow

will contain an average of all axes between the $\langle 1\bar{1}0 \rangle$ and $\langle 001 \rangle$ axes. The single-crystal elastic constants of tin³⁰ yield a compressional wave speed $v_{001} = 3.64 \text{ km s}^{-1}$ and $v_{110} (= v_{\bar{1}\bar{1}0}) = 3.18 \text{ km s}^{-1}$, with an average in the transverse plane of $v_{av} = 3.41 \text{ km s}^{-1}$. From the solidified ingot I machined a $20 \times 20 \times 40 \text{ mm}^3$ section from the columnar zone. Using the pulse and echo technique³¹ at three ultrasonic frequencies, I measured the compressional wave speed in the direction of heat flow ($v_{\parallel}$) and in each of two transverse directions ($v_{\perp 1}$ and $v_{\perp 2}$). The results are listed in Table 1. For all three frequencies the direction of heat flow and average dendritic growth has a speed $v_{\parallel}$ just slightly faster than v_{110} for a single crystal, whereas in the two transverse directions $v_{\perp 1}$ and $v_{\perp 2}$ are both quite

close to v_{av} . Table 1 shows that solidification texturing can produce the maximum achievable elastic anisotropy given the arbitrary orientation of the $\langle 1\bar{1}0 \rangle$ and $\langle 001 \rangle$ axes in directions perpendicular to that of dendritic growth of tin. Although lamination due to solvent-rich dendrites and solute-rich interdendritic material could give rise to an elastic wave speed anisotropy, in both the inner core and in the laboratory experiments the fraction of interdendritic material is low enough to make this an unlikely source of the anisotropy.

An obvious difficulty in using solidification texturing to explain the observed seismic anisotropy is that one would expect spherically symmetric cooling to yield a radial/transverse anisotropy, not a

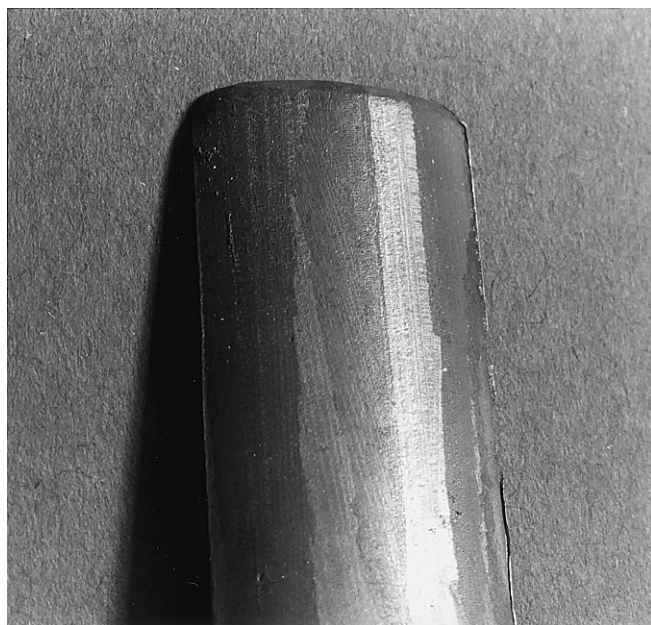


Figure 1 Photograph of the exterior surface of a slice of a directionally solidified, 97% Sn–3% Pb (by weight) alloy. The 400-g cylindrical ingot is 35 mm in diameter and ~55 mm tall. The melt was cooled from the bottom in a fixed furnace at a rate of 2.5 K min^{-1} with a maximum temperature gradient of 0.8 K mm^{-1} . Etching with 85% H_2O , 5% HNO_3 and 10% HCl revealed good contrast between tin crystals. The chill zone, consisting of small, equiaxed crystals, is at the ingot bottom (beneath the field of vision). Columnar crystals elongated in the direction of heat flow grow out of those crystals in the chill zone that have the most favourable lattice orientation with respect to the direction of heat flow, thereby allowing primary dendrites to grow rapidly⁸. Within each crystal the primary dendrite arms appear as a series of long parallel lines (they must be parallel to preserve the crystal lattice). Many primary dendrites can be seen within each crystal, so that a mushy zone can exist within a single crystal. From one crystal to another the dendrites need not be exactly parallel, though they are generally in the direction of heat flow. The arms are visible on the ingot exterior surface because of solidification shrinkage. The shorter secondary arms, perpendicular to the primary arms, are also apparent. By polishing and etching a longitudinal interior section, the lead is revealed as divorced eutectic along the dendrite arms (tin permits very little lead into its lattice). Thus, within the larger tin crystals (typically a few millimetres in width and nearly the same height as the ingot itself), lead forms $\sim 10^{-2} \text{ mm}$ crystallites.

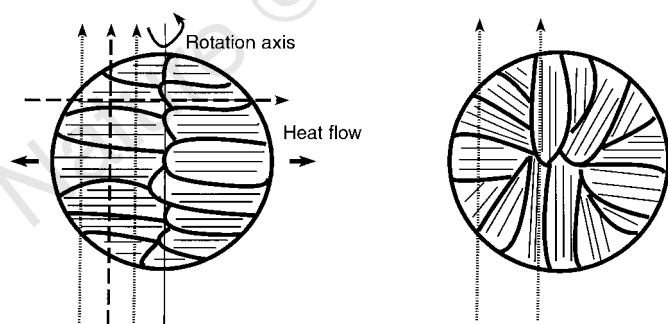


Figure 2 Left, diagram of the Earth's inner core in a plane containing the rotation axis, showing the direction of heat flow. Right, diagram showing the equatorial plane of the inner core. Heavy lines represent columnar crystals composed of parallel primary dendrites, which are represented by light lines. Seismic rays turning at different depths in the equatorial plane are shown by the dotted lines, those lying in a plane containing the rotation axis by dashed lines (rays parallel to the rotation axis both lie in a plane containing the rotation axis and turn in the equatorial plane). Rays parallel to the rotation axis and turning at any depth propagate perpendicular to primary dendrites, so that they lie in a direction with a range of speeds between slow and fast because the crystals have different transverse orientations. In the equatorial plane rays turning at the centre of the inner core always lie in the primary dendritic, or slow, direction. On the other hand, rays turning just beneath the inner-core surface tend to be perpendicular to primary dendrites, so that these rays, like rays parallel to the rotation axis, traverse an ensemble of crystals with different transverse orientations. Hence, at the inner-core surface the anisotropy disappears for rays with turning points in the equatorial plane. At the other extreme, rays in a plane containing the rotation axis will generally show a greater anisotropy that does not vanish at the inner-core surface. Also, columnar crystals elongated in the cylindrically radial direction may be responsible for an attenuation anisotropy, with a greater attenuation for compressional waves propagating parallel to the rotation axis. For a given turning radius, such waves cross more crystal boundaries than do waves propagating perpendicular to the rotation axis. In the short-wavelength limit, crystal boundaries cause attenuation (resulting in a decrease in the transmitted amplitude) due to an impedance contrast^{32,33} caused by the arbitrary rotation of each crystal about the dendritic direction. In the long-wavelength limit, crystal boundaries result in Rayleigh scattering. In either case one might expect an attenuation anisotropy, which should show a ray path dependence similar to that for the elastic anisotropy.

cylindrical anisotropy. But convection in a rapidly rotating spherical shell may be more efficient at withdrawing heat and light solute in equatorial regions than nearer the poles, even when a magnetic field is present¹⁰. This could lead to cylindrically radial heat flow near the inner–outer core boundary^{16,17}, with preferential growth of the inner core in the equatorial zone. In this case one would expect

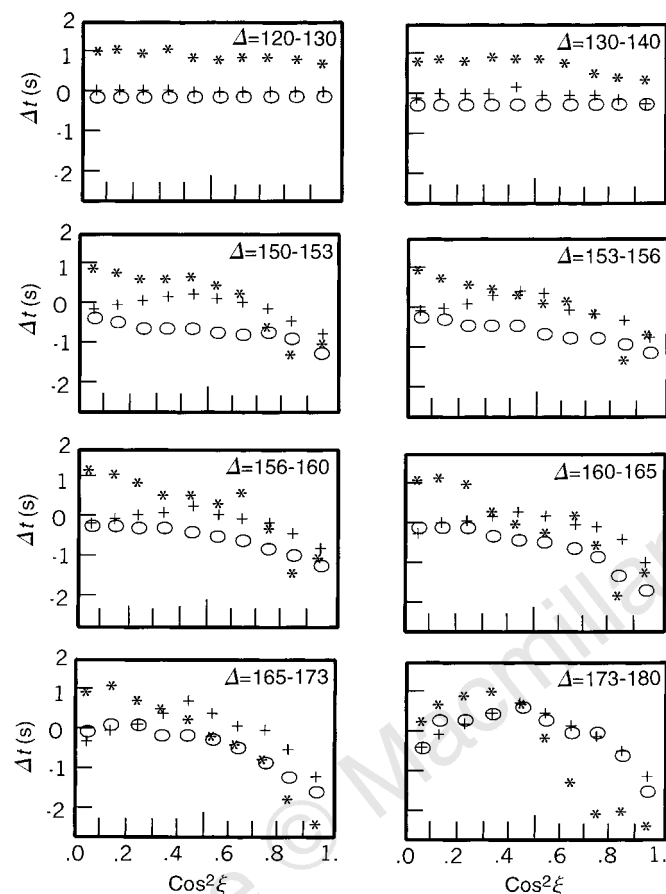


Figure 3 Compressional wave travel-time residuals inferred seismologically (asterisks)* and predicted by solidification texturing for h.c.p. iron (I show two extremes: ellipses for rays turning in the equatorial plane, plus-signs for rays in a plane containing the rotation axis) as a function of ray angle ξ from the rotation axis, for eight surface epicentral distance ranges Δ . The seismological residuals have been computed using the rotation axis as the axis of symmetry. The 1-s offset for rays turning in the shallow inner core ($\Delta = 120\text{--}130^\circ$) is simply due to the reference Earth model used, and is an indication that in comparing the seismological and predicted residuals a uniform shift is possible. To compute the predicted travel time as a function of ξ and Δ I integrated along the (assumed) straight ray path, using a local wave speed determined by the minimum possible angle α between the ray and the c -axis, as dictated by the solidification model, at each point along its path. For rays turning in the equatorial plane $\alpha = \sin^{-1}(\sin \xi \cos \phi)$, where ϕ is the angle in the equatorial plane between a fixed radius parallel to the ray and the radius through each point along the ray path. For rays in a plane containing the rotation axis, $\phi = 0$, or $\alpha = \xi$. Because of the possible rotation of crystals about the $\langle 210 \rangle$ axis, the local wave speed is then the average of the theoretical values in the $\langle 001 \rangle - \langle 010 \rangle$ plane¹⁵, between α and $\pi/2$. For example, for a ray with $\Delta = 180^\circ$ ($\phi = 0$) and $\xi = 0$ the possible wave speeds will range between those of $\langle 001 \rangle$ and $\langle 010 \rangle$, but for a ray with $\Delta = 180^\circ$ and $\xi = \pi/4$ the possible wave speeds will range only between those of $\langle 011 \rangle$ and $\langle 010 \rangle$. As the c -axis is fast, the former ray will produce a smaller travel time. I then subtract the travel time for a Voigt–Reuss–Hill isotropic inner core to yield the plotted predicted residual. The minimum in the theoretical compressional wave speed is between the $\langle 011 \rangle$ and $\langle 010 \rangle$ axes, and is responsible for the increase in some residuals at intermediate ξ .

primary dendrites pointed cylindrically outward to outgrow others, and the crystals that comprise these dendrites to be elongated in the cylindrically radial direction (Fig. 2). This preferential growth in equatorial regions may be ultimately inhibited by the decrease in pressure with radius limiting the solidification (in spite of the cooler temperature and lower solute concentration due to more effective transport). Alternatively, the inner core may adjust isostatically by means of solid-state flow that could result in a deformation or recrystallization texture¹⁷. However, the timescale for the development of such a texture is long, comparable with the age of the Earth¹⁷. Moreover, such a texture must develop from the solidification texture already present, and the process of recovery can preserve the original texture^{19,20}.

The phase of iron in the inner core is not known with certainty. However, the elastic constants of h.c.p. and f.c.c. iron at core pressure have been estimated theoretically¹⁵. For f.c.c. iron the axis of fastest compressional wave speed is $\langle 111 \rangle$ and for h.c.p. iron it is $\langle 001 \rangle$, the c -axis. In either case, the direction of dendritic growth is not a fast axis, consistent with cylindrically radial growth of dendrites and the direction parallel to the rotation axis being seismically fast. The solidification-texturing mechanism dictates that for h.c.p. crystals the cylindrically radial direction will contain $\langle 210 \rangle$ crystal axes. For a given crystal there is no preferred direction in the plane transverse to that of dendritic growth, so that the $\langle 010 \rangle$ axis (which is elastically equivalent to the $\langle 210 \rangle$ axis) and the $\langle 001 \rangle$ axis can lie in any perpendicular directions in the plane locally tangent to a cylinder about the rotation axis.

Applying the theoretical values of the h.c.p. iron elastic constants to the distribution of crystal orientations predicted by cylindrically radial solidification texturing, I computed the compressional wave travel-time residuals as a function of epicentral distance and ray angle from the rotation axis (Fig. 3). The simple physical model reproduces the general features of the observed residuals, although with a maximum predicted residual of only about 1–1.5% in the well-sampled $150\text{--}153^\circ$ range of epicentral distance. The small anisotropy of the model arises because of the averaging of crystal axes in the plane transverse to the dendritic growth direction. However, the zero-temperature theoretical elastic constants¹⁵ could underestimate the single-crystal anisotropy at higher temperatures. For example, α -titanium and α -zirconium at atmospheric pressure often serve as analogues for h.c.p. iron¹³, and in each of these v_{100}/v_{001} is observed to decrease continuously with temperature, from 0.96 at 4 K to 0.90 at $0.54 T_m$ (where T_m is the melting temperature) for α -titanium, and from 0.95 at 4 K to 0.90 at $0.41 T_m$ for α -zirconium³⁰. Note that solidification texturing can

Table 1 Measured wave speeds

	Frequency (MHz)		
	2.25	5.0	10.0
v_l (km s ⁻¹)	3.30	3.23	3.25
$v_{\perp 1}$ (km s ⁻¹)	3.35	3.38	3.40
$v_{\perp 2}$ (km s ⁻¹)	3.40	3.40	3.40

Compressional elastic wave speeds in a 20 mm wide by 20 mm wide by 40 mm high right parallelepiped section of tin alloy cut from above the lowest 7.5 mm of the ingot to avoid the equilaxed chill zone. At 2.25 and 5.0 MHz I used 12.5-mm-diameter transducers, and at 10.0 MHz a 6.25-mm-diameter transducer. The speeds $v_{\perp 1}$ and $v_{\perp 2}$ are transverse to the direction of heat flow; v_l is in the direction of heat flow. The difference between the speeds in two transverse directions versus that in the parallel direction is evident. The anisotropic speeds agree with that predicted by solidification texturing (see text). At these frequencies, the wavelength in tin is in the range 1.5–0.3 mm, a small fraction of the sample size and of the columnar crystal height, but a sizeable fraction of the typical crystal width of a few millimetres (see Fig. 1). Under the laboratory solidification conditions a columnar crystal typically comprises of the order of ten primary dendrites. A crude extrapolation to planetary cores, given by multiplying the meteoritic primary arm spacing of tens of metres (ref. 23) by the typical number of dendrites in a columnar crystal, suggests a characteristic inner-core crystal width of hundreds of metres. If the inner-core crystals are columnar in shape due to directional cooling then their height may be very much longer than hundreds of metres. Thus in the inner core, the typical seismic body wave may be sampling crystals with a length scale that in one direction is longer than the wavelength (about 10 km at 1 Hz), but in other directions is smaller than (though perhaps a significant fraction of) the wavelength.

result in an increase in compressional wave speed anisotropy with turning depth, which can be seen from Fig. 2: this figure also shows how columnar crystals elongated in the cylindrically radial direction could be the cause of an attenuation anisotropy.

Solidification texturing, and columnar crystals due to the preferential extraction of heat and solute from the inner core in the cylindrically radial direction, seem broadly capable of causing the inner-core elastic anisotropy, its depth dependence, and an attenuation anisotropy. Uncertainty remains whether, owing to the necessary averaging in the transverse directions, the magnitude of the elastic anisotropy developed by solidification texturing is sufficient to match the seismic observations, and whether possible subsequent solid-state flow modifies the solidification texturing. It is also not clear why the symmetry axis is tilted from the rotation axis by a few degrees. Further work in seismology, high-pressure mineral physics, magnetohydrodynamics and solidification should elucidate this most remote region of the Earth. □

Received 25 November 1996; accepted 21 July 1997.

- Morelli, A. D., Dziewonski, A. M. & Woodhouse, J. H. Anisotropy of the inner core inferred from PKIKP travel times. *Geophys. Res. Lett.* **13**, 1545–1548 (1986).
- Woodhouse, J. H., Giardini, D. & Li, X. D. Evidence for inner core anisotropy from free oscillations. *Geophys. Res. Lett.* **13**, 1549–1552 (1986).
- Creager, K. C. Anisotropy of the inner core from differential travel times of the phases PKP and PKIKP. *Nature* **356**, 309–314 (1992).
- Tromp, J. Support for anisotropy of the Earth's inner core from free oscillations. *Nature* **366**, 678–681 (1993).
- Shearer, P. Constraints on inner core anisotropy from PKP(DF) travel times. *J. Geophys. Res.* **99**, 19647–19659 (1994).
- Su, W. & Dziewonski, A. M. Inner core anisotropy in three dimensions. *J. Geophys. Res.* **100**, 9831–9852 (1995).
- Song, X. Anisotropy in central part of inner core. *J. Geophys. Res.* **101**, 16089–16097 (1996).
- Porter, D. A. & Easterling, K. E. *Phase Transformations in Metals and Alloys* (Chapman & Hall, London, 1992).
- Chalmers, B., *Principles of Solidification* (Wiley, New York, 1964).
- Glatzmaier, G. A. & Roberts, P. H. Dynamo theory then and now. *Int. J. Eng. Sci.* (in the press).
- Cornier, V. F. Inner core structure inferred from body waveforms. *Eos* **75**, 67 (1994).
- Souriau, A. & Romanowicz, B. Anisotropy in inner core attenuation: a new type of data to constrain the nature of the solid core. *Geophys. Res. Lett.* **23**, 1–4 (1996).
- Wenk, H. R., Takeshita, T., Jeanloz, R. & Johnson, G. C. Development of texture and elastic anisotropy during deformation of hcp metals. *Geophys. Res. Lett.* **15**, 76–79 (1988).
- Karato, S. I. Inner core anisotropy due to the magnetic field-induced preferred orientation of iron. *Science* **262**, 1708–1711 (1993).
- Stixrude, L. & Cohen, R. E. High-pressure elasticity of iron and anisotropy of Earth's inner core. *Science* **267**, 1972–1975 (1995).
- Stacey, F. *Physics of the Earth* (Brookfield, Brisbane, 1992).
- Yoshida, S., Sumita, I. & Kumazawa, M. Growth model of the inner core coupled with the outer core dynamics and the resulting elastic anisotropy. *J. Geophys. Res.* **101**, 28085–28103 (1996).
- Rutter, J. W. in *Liquid Metals and Solidification* (ed. Maddin, R.) 243–262 (American Society of Metals, Cleveland, 1958).
- Haasen, P. *Physical Metallurgy* (Cambridge Univ. Press, 1978).
- Barrett, C. & Massalski, T. B. *Structure of Metals* (Pergamon, New York, 1980).
- Sherman, D. M. Stability of possible Fe–FeS and Fe–FeO alloy phases at high pressure and the composition of the Earth's core. *Earth Planet. Sci. Lett.* **132**, 87–98 (1995).
- Fearn, D. R., Loper, D. E. & Roberts, P. H. Structure of the Earth's inner core. *Nature* **292**, 232–233 (1981).
- Esbensen, K. H. & Buchwald, V. F. Planet(oid) core crystallization and fractionation—evidence from the Apaglik mass of the Cape York iron meteorite shower. *Phys. Earth. Planet. Inter.* **29**, 218–232 (1982).
- Bergman, M. I., Fearn, D. R., Bloxham, J. & Shannon, M. Convection and channel formation in solidifying Pb–Sn alloys. *Metallurgical Transactions A* **28**, 859–866 (1997).
- Bergman, M. I. & Fearn, D. R. Chimneys on the Earth's inner–outer boundary? *Geophys. Res. Lett.* **21**, 477–480 (1994).
- Worster, M. G. Natural convection in a mushy layer. *J. Fluid Mech.* **224**, 335–359 (1991).
- Loper, D. E. Structure of the inner core boundary. *Geophys. Astrophys. Fluid Dyn.* **25**, 139–155 (1983).
- Doombos, D. J. The anelasticity of the inner core. *Geophys. J. R. Astron. Soc.* **38**, 397–415 (1974).
- Saxena, S. K. *et al.* Synchrotron X-ray study of iron at high pressure and temperature. *Science* **269**, 1703–1704 (1995).
- Simmons, G. & Wang, H. *Single Crystal Elastic Constants and Calculated Aggregate Properties* (MIT Press, Cambridge, MA, 1971).
- McSkimin, H. J. Ultrasonic measurement techniques applicable to small solid specimens. *J. Acoust. Soc. Am.* **22**, 413–418 (1950).
- Roth, W. Scattering of ultrasonic radiation in polycrystalline metals. *J. Appl. Phys.* **19**, 901–910 (1948).
- Mason, W. P. & McSkimin, H. J. Energy losses of sound waves in metals due to scattering and diffusion. *J. Appl. Phys.* **19**, 940–946 (1948).

Acknowledgements. I thank K. Bishop (Matec Instruments) and B. Li for assistance with the ultrasonic measurements; J. Bloxham, W. Dunbar, D. Fearn, C. Francis, W. Kuang, H. Ma, R. O'Connell, M. Shannon, F. Spaepen, J. Tromp, and especially S. Zatman for suggestions; G. Glatzmaier and S. Yoshida for providing preprints; and J. Bloxham, supported by the David and Lucile Packard Foundation and the NSF, and Harvard University, for providing facilities. This work was supported by a Cottrell College Science Award of the Research Corporation.

Correspondence and requests for materials should be addressed to the author (e-mail: bergman@simons-rock.edu).

An intrinsic frequency limit to the cochlear amplifier

Jonathan E. Gale*† & Jonathan F. Ashmore*‡

* Department of Physiology, School of Medical Sciences, University Walk, Bristol BS8 1TD, UK

‡ Department of Physiology, University College London, Gower Street, London WC1E 6BT, UK

Hearing in mammals depends on a feedback process within the inner ear termed the 'cochlear amplifier'. The essential components of this amplifier are sensorimotor cells, the outer hair cells, which transduce motion of the basilar membrane induced by sound and generate forces to cancel the viscous damping of the cochlear partition^{2,3}. Outer hair cells alter the passive mechanics of the cochlea, enhancing both the sensitivity and the frequency selectivity of the auditory system. The molecular basis of the mechanism is thought to be a voltage-sensitive 'motor' protein, as yet unidentified, embedded in the basolateral membrane of the outer hair cell⁴. The cochlear amplifier operates up to at least 22 kHz (ref. 5), but by measuring both the charge and mechanical movements associated with the motor^{6,7} in isolated membrane patches under voltage clamp, we show here that the limiting frequency at which the motor operates lies near 25 kHz. This value therefore sets an upper limit to the range of hearing in mammalian cochleas using this mechanism. The fast charge movement, arising from charge displacement within the presumed motor molecule, further suggests that the protein is more likely to be related to a transporter than to a modified ion channel.

The lateral membrane of an outer hair cell (OHC) contains a sufficient density of membrane motor particles^{4,8} for single patches to produce a measurable charge movement without ionic current flow when the membrane potential is changed. To obtain a greatly enhanced temporal resolution, we made recordings in cell-attached and excised patches using low access-resistance (1–4 MΩ) pipettes⁹. Transient currents, largely independent of ionic currents and resembling gating currents of ion channels, were obtained using standard techniques¹⁰ (Fig. 1a). The charge *Q* moved across the membrane during voltage steps was measured between –160 mV and +100 mV, and the voltage dependence was described (Fig. 1d) by

$$Q(V) = Q_{\max} / [1 + \exp(-\beta(V - V_0))]$$

which is a Boltzmann distribution where $Q_{\max}$ represents the maximum charge transferred, V_0 is the potential at which the charge is distributed equally^{6,7,11}, $\beta = ze\delta/k_B T$ and is a constant given by the number, *z*, of elementary charges, *e*, displaced across a fraction, δ , of the membrane dielectric, k_B is Boltzmann's constant and *T* is the absolute temperature. This model describes the movement of positive charge from the inside to the outside of the membrane when the membrane potential is stepped from a negative to a positive potential. Wide-band recordings (Fig. 1a, b) indicate that the transient current relaxation could be fitted to a single exponential with a time constant of between 7 and 50 μs. The values depend upon the final potential and so different time constants were obtained for 'on' and 'off' voltage steps within the same pulse, showing that the responses were not limited by recording bandwidth. The mean time constant averaged for steps to –20 mV was $13.6 \pm 6.4 \mu s$ (mean $\pm$ s.d., *n* = 10). The shortest time constants measured in 10 patches were $9.9 \pm 1.4 \mu s$ for 'on' steps to depolarized potentials and $10.5 \pm 3.0 \mu s$ for 'off' steps to hyperpolarized

† Present address: Departments of Otolaryngology and Neuroscience, University of Virginia, Charlottesville, Virginia 22908, USA.

potentials. In 50% of patches the best fit to the transient current was obtained using a combination of two exponentials. The second exponential had a significantly slower time course (range 23 μ s to 2.3 ms), which we ascribe to residual contamination by ionic currents. The time constants were temperature dependent (Fig. 1c) with $Q_{10} = 1.34$ in the range 10–26 °C. Bath application of 10 mM sodium salicylate reduced the charge movement to $28.5 \pm 14.3\%$ of control ($n = 3$; Fig. 1e). Salicylate also reduces transient currents and motility in whole-cell recording¹² and provides further evidence that the transient currents in patches arise from the motor protein.

Voltage steps caused the membrane patch in the pipette to move⁴. The movement was detected optically using a focused laser beam, and the time course of the patch displacement followed $Q(t)$ closely with no measurable delay in six out of seven patches (Fig. 2), although the rise time for patch movement was greater. Patch displacement was fitted with a single time constant, $83.5 \pm 22.3 \mu$ s

($n = 7$, range 7–162 μ s, the extremes being for excised patches). Discrepancies between the time constants for charge and mechanical movements have also been noted in whole-cell recordings^{13,14}, where the likely cause is the internal viscous damping of the cell. In the present case, single patches would be constrained by the cortical cytoskeleton and would be susceptible to viscous damping when a large column of fluid is moved within the pipette tip. The change in the radius of the patch was estimated to be 0.05 nm mV^{-1} , and implies that the area increased by 0.8% for a 100-mV step stimulus, assuming that the patch is a hemisphere of diameter 2.5 μ m. There was no patch movement or transient currents when recordings were made in cochlear supporting cells.

As a charge Q moving across the membrane generates an apparent voltage-dependent membrane capacitance, $C_V = dQ/dV$, we used the magnitude of the capacitance measured with sinusoidal commands to estimate the limiting kinetics of the motor in the frequency domain¹⁵. Capacitance was measured in 19 patches. As

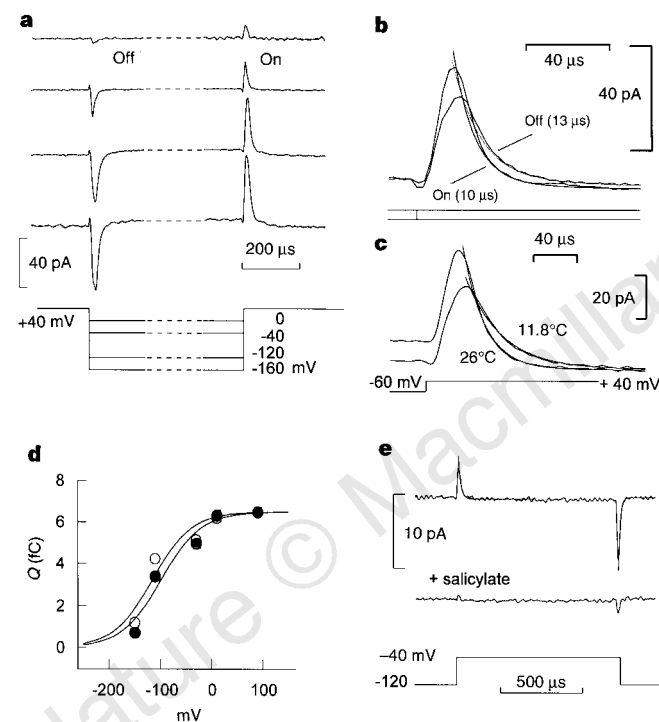


Figure 1 Rapid charge movement associated with the OHC motor in isolated membrane patches. **a**, Charge movement measured as transient currents. From top to bottom, currents were recorded, from +40 mV holding potential in a cell-attached patch, with steps to +0, -40, -120 and -160 mV. 600 records were averaged with a P/4 protocol¹⁴ and smoothed over four points for display. Sample rate, 333 kHz, low-pass filtered at 60 kHz. The currents were fitted with two exponentials, the faster exponential associated with the 'motor'. A slower current, likely to be the residual macroscopic current resulting from channel activity, was not present in all patches. The shortest fitted time constant varied between 7 and 50 μ s. **b**, The 'on' (to +40 mV) and 'off' (from +40 mV) charge movements are shown for the step to -120 mV on an expanded time scale for comparison. These data were best fitted with two exponentials (in μ s): on (off) τ_1 10 (13), τ_2 142 (205). **c**, The time constants decreased with bath temperature. The on response for a cell-attached patch shown with single exponential fits to the data. Q_{10} values for on and off time constants were equal. The total charge transferred was independent of temperature. Cell-attached patch, low-pass filtered at 10 kHz, and averaged 1,000 times with a P/4 protocol. The baseline at 26 °C was due to unsubtracted ionic conductance. **d**, Charge movement (from **a**) as a function of voltage. Open circles, Q_{on} ; filled circles, Q_{off} . The fitted Boltzmann function has a midpoint at -104 (-88) mV, and $\beta = 0.066$ (0.065) mV^{-1} (equation (1) with $Q_{max} = 6.5 \text{ fC}$). The difference between the curves is statistically not significant. **e**, Bath application of 10 mM sodium salicylate reversibly blocked the charge movement. Single exponential fits with τ_{on} 11 μ s, τ_{off} 17 μ s.

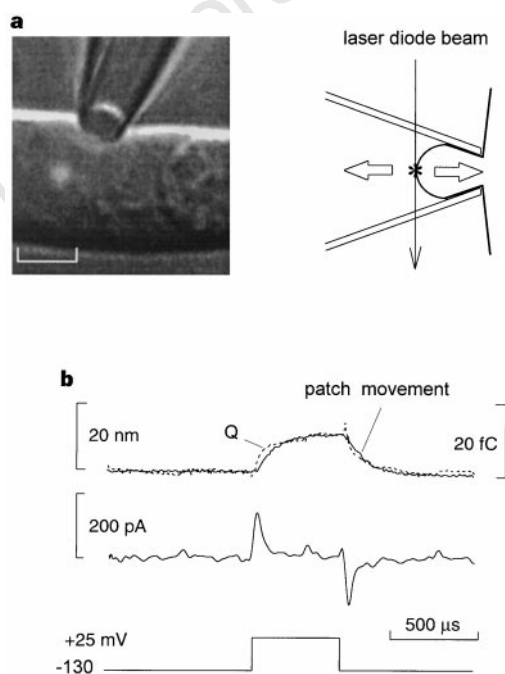


Figure 2 Movement of an isolated OHC patch. **a**, Diagram of the experimental arrangement, showing direction of patch movement. Left, patch of membrane in pipette recording from basolateral membrane of the cell. Scale bar, 5 μ m. Right, a diode laser beam was focused to a 2- μ m spot using a 40 $\times$ Wl objective as condenser and the spot scanned over the patch region to intersect the membrane interface. Transmitted light was detected by a wide-band differential photodiode at the exit pupil, and averaged to give a motion signal. Movements were calibrated by displacing the whole pipette to a known distance. **b**, Simultaneous patch and charge movement. Membrane patches moved by up to 20 nm upon stepping the patch potential from -130 to +25 mV; the magnitude of the movement was graded with the size of the voltage step and saturated at hyperpolarized and depolarized potentials (measured using triangular wave-form command voltages; data not shown). Top, calibrated photodiode record (solid). Middle, transient current from P/4 protocol, averaged 3,000 times and smoothed for clarity. The integral of the current, Q , has been scaled and plotted over the photodiode record (dashed line, top trace). Bottom, voltage command. Single exponential time constants for the patch and charge movement are 103 and 22.5 for on, and 107 and 19.3, respectively, for off.

the frequency of the command was changed from 0.5 to 90 kHz, there was a reduction in the maximum value, $C_{\max}$, of C_V (Fig. 3). Such a reduction would arise if there were a finite rate for the translocation of charge with the consequence that $C_{\max}$ would become a function of frequency f given by

$$C_{\max}(f) = C_{\max}(0) / \sqrt{1 + (2\pi\tau f)^2}^{1/2}$$

where τ is the relaxation time constant of the charge movement; the corresponding frequency, giving the effective bandwidth, would therefore be $f_c = 1/2\pi\tau$. There are limitations to the bandwidth in patch recording, but in measurements of the voltage-dependent capacitance, the capacitance of the pipette does not restrict $C_{\max}$. The limiting frequency of the electronics, however, became significant above 60 kHz (ref. 16). A fit to the pooled data incorporating this filter produced an average f_c of 9.95 kHz (Fig. 3b), equivalent to a time constant τ of 16.0 μ s. If f_c was obtained by fitting C_V as a function of frequency for individual patches, the resultant mean f_c was 10.6 ± 1.3 kHz (s.e.m., $n = 19$). The fitting procedure to C_V produces a low value for f_c as it depends on a weighted average over the whole membrane voltage range. The voltage dependence of the capacitance became less pronounced at frequencies above 10 kHz, which is consistent with the corner frequency, f_c , being voltage dependent (as implied in Fig. 1). No voltage-dependent membrane capacitance was observed in patches taken from the basal pole of OHCs⁹ or from supporting cells.

The fits to these data strongly suggest that there is a limiting step in the operation of the OHC. The underlying movement of charge is faster than the gating charge movement described in ion channels¹⁷, where the spatial rearrangement of charged residues within the protein of a voltage-sensitive ion channel produces fluctuations in the macroscopic gating current. In single patches, dipole transitions can be subjected to fluctuation analysis¹⁸. However, application of this technique to the charge movements in OHCs, in both membrane patches and whole cells, failed to reveal any underlying elementary fluctuations (data not shown) indicating significant differences from ion channel behaviour.

The rapid charge movement in OHCs is as fast as that described in the operation of ion-transporting proteins¹⁹. Models for these proteins suggest that a significant fraction of the rapid charge movement may be carried by the first ion entering the pore or access channel spanning a large fraction of the membrane²⁰. To test the prediction that extracellular ions modify the charge movement process, we changed external ion concentration around the OHC while recording membrane capacitance under whole-cell recording conditions, and measured the shift in the potential, V_0 , at which the

capacitance was maximal. Reducing extracellular chloride from 160 to 5 mM by replacement with gluconate ($n = 4$) had no effect on capacitance or total current. This finding indicates that OHCs do not possess chloride channels, and suggests that the motor is unlikely to be a transporter specialized for anions, such as a member of the anion-exchange superfamily. Replacement of external sodium with sucrose solution reversibly shifted V_0 to negative potentials with a slope of 23 mV per decade ($n = 17$) in the range 160–20 mM, but replacement of sodium with either *N*-methyl-D-glucamine (NMG; $n = 6$) or with triethanolamine ($n = 3$) did not produce a significant shift in V_0 .

The data are compatible with a model in which the charge is carried by cations entering a variable-geometry access channel into which anions do not permeate²¹, although there are other hypotheses. The effects of lowered Na_o alone, without replacement by NMG, could be explained by charge screening on the external membrane; in this case, we calculate that the OHC would possess a density of 1 negative charge per 16 nm² to account for the observed shift in sucrose. Alternatively, the patch motion (Fig. 2b) suggests that otherwise hidden charge residues around the pore could be exposed to produce an apparent charge movement when the area increases, a mechanism that has been proposed for sodium-channel gating²². However, none of the hypotheses alone can explain all the data, and the most economical summary of the physical constraints on the motor molecule, especially its rate, points to a protein related to a transporter undergoing a conformational change.

These results set an upper limit to the hearing range of cochleas, which depend on cochlear amplification supplied by OHC force feedback. The data describe the limiting frequency at which the OHC motor protein can undergo conformational changes before generating forces in the cycle. However, the frequencies and rates measured at –20 mV should be corrected upwards to account for the negative resting potential of the cell *in vivo*. Taking the smallest mean 'on' time constant, 9.9 μ s, the equivalent corner frequency would be 16.1 kHz. Corrected to 37 °C using a Q_{10} of 1.33 (Fig. 1c)²³, this intrinsic limit would be 26 kHz. It may be no coincidence that, with the exception of cochleas specialized to detect ultrasound, the upper limit to the auditory range of most terrestrial mammals is within a factor of two of this limiting frequency. The enhanced auditory range in specialized echo-locating species may therefore arise through further modified mechanisms, for these species have greatly different cochlear structures²⁴. It has also been suggested that *in vivo* there may be mechanisms to extend the frequency response of cochlear amplifier in nonspecialized mammalian cochleas⁵. Nevertheless, our results indicate that there is an intrinsic time constant associated with the conformational change of the OHC

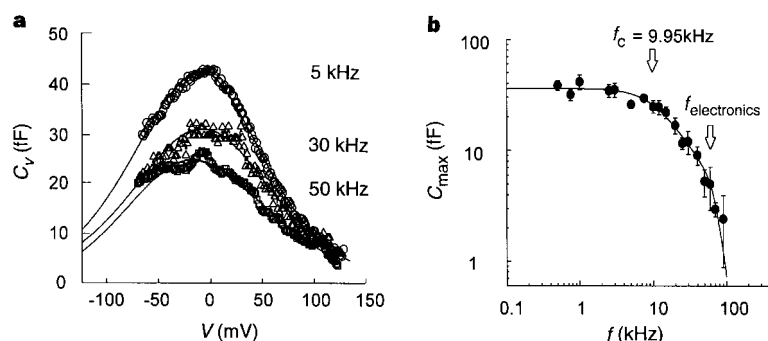


Figure 3 Voltage-dependent capacitance in isolated membrane patches. **a**, Membrane capacitance C_V plotted as a function of membrane potential using command frequencies of 5, 30 and 50 kHz. Each symbol represents an individual measurement of capacitance made at a rate of 10 Hz during a voltage ramp from –80 to +130 mV. Increasing the frequency of the sinusoidal command signal decreased the maximum values of C_V . **b**, Pooled data from 17 patches from different cells. Filled circles represent the mean $\pm$ s.e.m. of the maximum

voltage-dependent capacitance recorded from 4–19 patches as a function of the sinusoid frequency. Not all frequency points were obtained in all patches. The solid line is given as equation (2), with the lowest frequency cut-off at $f_c = 9.95$ kHz, equivalent to a time constant τ of 17.0 μ s (arrow marked f_c). The time constant of the amplifier electronics (2.7 μ s or equivalently 60 kHz) is limiting only at high frequencies (shown by arrow marked $f_{\text{electronics}}$).

motor, preceding the force generation, that produces a cut-off frequency above which power cannot effectively be injected into cochlear mechanics. □

Methods

Cell preparation. The isolation technique was similar to that described previously^{12,13}. Adult guinea-pigs (200–400g) were killed by rapid cervical dislocation and both bullae were removed. The organ of Corti was dissected in a phosphate-buffered balanced salt solution containing (mM): 142, NaCl; 4, KCl; 1, CaCl₂; 1.5, MgCl₂; 8, Na₂HPO₄; 2, NaH₂PO₄ adjusted to pH 7.25. The osmolality of the solution was adjusted to 323–325 mosM using D-glucose. Cells were isolated by triturating the dissected tissue after incubation in 0.25 mg trypsin per ml (Sigma, type III) for 15 min. Extracellular solutions contained 142 NaCl, 4 KCl, 1.5 MgCl₂, 1 CaCl₂, 10 HEPES, adjusted to pH 7.3 with 1 M NaOH and to 325 mOsm using D-glucose. Pipette solutions for isolated patch experiments contained external solution with the addition of 10 mM BaCl₂ to block potassium channels. Conventional patch and whole-cell voltage-clamp recording techniques used an Axopatch 200 A amplifier (Axon, USA) to record from cells in a temperature-controlled chamber. Giga-seals were formed on the lateral walls of cells using low-resistance (1–4 MΩ) patch electrodes with the largest diameter possible (approximately 3 μm). Pipettes were coated with ski wax (Toko, Switzerland) or Sylgard to minimize capacitance. Patches were held at –20 mV to monitor the patch current, usually <1 pA. Data were filtered at 60 kHz and sampled at 333 kHz, unless stated. This sampling rate limited all time-resolved data to 3 μs. Standard P/4 methods were used to remove residual patch leak and linear capacitive currents¹⁰ based on evidence from whole-cell recording that the motor charge movement takes place over a limited voltage range. Patches were stepped from the holding potential to the leak holding potential of –130 or +40 mV, and data were averaged as indicated. Data were fitted using a Levenberg–Marquardt algorithm based on 250 points. The standard error of the time constant estimate was less than 1.2 μs.

Capacitance. Patch capacitance was measured using a conventional lock-in amplifier (Stanford Research SR 530) and a patch amplifier (Axopatch 200 A). Sinusoidal commands (10 mV below 10 kHz and 0.5 mV above 10 kHz to avoid amplifier saturation) were summed with a command ramp (–80 to +130 mV) and fed directly to the head stage. Capacitance at each frequency was calibrated using a 100 fF dither circuit. The phasor angle was adjusted to measure only the absolute change in capacitance. Measurements between 1 and 90 kHz were made in a pseudo-random order. The frequency-dependent capacitance of the OHC patch, $C(f)$, is a sum of a voltage-independent linear term C_{patch} and a voltage-dependent term $C_v(f)$. The latter was identified by using a Levenberg–Marquardt fitting algorithm to identify the derivative of the Boltzmann curve (equation (1)) at each frequency.

Received 1 May; accepted 10 June 1997.

1. Davis, H. An active process in cochlear mechanics. *Hearing Res.* **9**, 79–90 (1983).
2. Dallos, P. In *The Cochlea* (eds Dallos, P., Popper, A. N. & Fay, R. R.) 1–43 (Springer, New York, 1996).
3. Mammano, F. & Nobili, R. Biophysics of the cochlea: linear approximation. *J. Acoust. Soc. Am.* **93**, 3320–3332 (1993).
4. Kalinec, F., Holley, M. C., Iwasa, K. H., Lim, D. J. & Kachar, B. A membrane-based force generation in auditory sensory cells. *Proc. Natl Acad. Sci. USA* **89**, 8671–8675 (1992).
5. Dallos, P. & Evans, B. High frequency motility of outer hair cells and the cochlear amplifier. *Science* **267**, 2006–2009 (1995).
6. Ashmore, J. F. Forward and reverse transduction in the mammalian cochlea. *Neurosci. Res.* **11** (suppl.), 39–50 (1990).
7. Santos-Sacchi, J. Reversible inhibition of voltage dependent outer hair cell motility and capacitance. *J. Neurosci.* **9**, 2954–2962 (1991).
8. Forge, A. Structural features of the lateral walls in mammalian cochlear outer hair cells. *Cell Tiss. Res.* **265**, 473–483 (1991).
9. Gale, J. E. & Ashmore, J. F. The outer hair cell motor in membrane patches. *Eur. J. Physiol.* **434**, 267–271 (1997).
10. Armstrong, C. M. & Bezanilla, F. Inactivation of the sodium channel. II Gating current experiments. *J. Gen. Physiol.* **70**, 567–590 (1977).
11. Iwasa, K. H. A membrane motor model for the fast motility of the outer hair cell. *J. Acoust. Soc. Am.* **96**, 2216–2224 (1994).
12. Tunstall, M. J., Gale, J. E. & Ashmore, J. F. Action of salicylate on membrane capacitance of outer hair cells from the guinea pig cochlea. *J. Physiol. (Lond.)* **485**, 739–752 (1995).
13. Ashmore, J. F. A fast motile response in guinea pig outer hair cells: the cellular basis of the cochlear amplifier. *J. Physiol. (Lond.)* **388**, 323–347 (1987).
14. Santos-Sacchi, J. On the frequency limit and phase of outer hair cell motility: the effects of membrane filter. *J. Neurosci.* **12**, 1906–1916 (1992).
15. Neher, E. & Marty, A. Discrete changes of cell membrane capacitance observed under conditions of enhanced secretion in bovine adrenal chromaffin cells. *Proc. Natl Acad. Sci. USA* **79**, 6712–6716 (1982).
16. Maconochie, D. J., Fletcher, G. H. & Steinbach, J. H. The conductance of the muscle nicotinic receptor channel changes rapidly upon gating. *Biophys. J.* **68**, 483–490 (1995).

17. Sigg, D., Stefani, E. & Bezanilla, F. Gating current noise produced by elementary transitions in Shaker potassium channels. *Science* **264**, 578–582 (1994).
18. Conti, F. & Stühmer, W. Quantal charge redistributions accompanying the structural transitions of sodium channels. *Eur. Biophys. J.* **17**, 53–59 (1989).
19. Hilgemann, D. W. Channel-like function of the Na, K pump probed at microsecond resolution in giant membrane patches. *Science* **263**, 1429–1432 (1994).
20. Gadsby, D. C., Rakowski, R. F. & De Weer, P. Extracellular access to the Na,K pump: pathway similar to ion channel. *Science* **260**, 100–103 (1993).
21. Rakowski, R. F. Charge movement by the Na,K pump in *Xenopus* oocytes. *J. Gen. Physiol.* **101**, 117–144 (1993).
22. Yang, N. B., George, A. L. Jr & Horn, R. Molecular basis of charge movement in voltage gated sodium channels. *Neuron* **15**, 213–218 (1995).
23. Ashmore, J. F. & Holley, M. C. Temperature dependence of a fast motile response in isolated outer hair cells of the guinea-pig cochlea. *Q. J. Exp. Physiol.* **73**, 143–145 (1988).
24. Vater, M. & Lenoir, M. Ultrastructure of the Horseshoe bat's organ of Corti. I. Scanning electron microscopy. *J. Comp. Neurol.* **318**, 367–379 (1992).

Acknowledgements. This work was supported by the Hearing Research Trust, the Colt Foundation, the Royal Society, and the Wellcome Trust.

Correspondence and requests for materials should be addressed to J.F.A. (e-mail: j-ashmore@ucl.ac.uk).

Modulation of neuronal activity by target uncertainty

Michele A. Basso & Robert H. Wurtz

Laboratory of Sensorimotor Research, National Eye Institute, Bethesda, Maryland 20892-4435, USA

Visual scenes are composed of many elements and although we can appreciate a scene as a whole, we can only move our eyes to one element of the scene at a time. As visual scenes become more complex, the number of potential targets in the scene increases, and the uncertainty that any particular one will be selected for an eye movement also increases. How motor systems accommodate this target uncertainty remains unknown. The activities of neurons in both the cerebral cortex^{1–5} and superior colliculus^{6–8} are modulated by this selection process. We reasoned that activity associated with target uncertainty should be evident in the saccadic motor system at the final stages of neural processing, in the superior colliculus^{9,10}. By systematically changing the number of stimuli from which a selection must be made and recording from superior colliculus neurons, we found that as the target uncertainty increased, the neural activity preceding target selection decreased. These results indicate that neurons within the final common pathway for movement generation are active well in advance of the selection of a particular movement. This early activity varies with the probability that a particular movement will be selected.

The intermediate layers of the superior colliculus receive input from extensive regions of the cerebral cortex and may be regarded as a virtual readout of the cerebral cortex for the control of eye movements. Moreover, prelude burst neurons and buildup neurons in the superior colliculus increase their activity with the selection and preparation of a movement long before the activity closely related to saccade generation^{6–8}. We determined the effect of target uncertainty on buildup neurons in awake monkeys in two ways. In the first experiment, we manipulated target uncertainty by presenting a different number of possible targets to monkeys on different trials. In a second experiment, we manipulated target uncertainty by presenting the same visual array of possible targets on every trial and varying the probability that any one of the stimuli would be selected for a saccadic eye movement. In both experiments we found that buildup neurons decreased their activity as uncertainty increased.

One, two, four or eight possible targets were introduced to increase uncertainty (Fig. 1: array on, uncertainty period), and then later the monkeys were cued as to which target was relevant (target dim, selection period). With a single possible saccadic target (Fig. 2, top row), neurons showed a transient visual response (left

spike density function) a subsequent sustained response which continued after the target dimmed (centre column), and then a burst of activity before the saccade to that target (right column).

During the period of uncertainty, increasing the number of possible targets in the array from one to two, to four and then to eight (Fig. 2) reduced the initial visual response. The increasing number of possible targets also decreased the activity after this initial visual response. At the end of the uncertainty period (before the dimming) the reduction continued as emphasized by the bold outline traces in the middle column of Fig. 2.

When the target dimmed, the uncertainty dropped to zero. In the single possible target case (Fig. 2, top row, middle column) there was no target uncertainty, the firing rate was already high and did not increase much further. With eight possible targets (bottom row, middle histogram), when there was maximal uncertainty in this task, the firing rate increased when the target was selected.

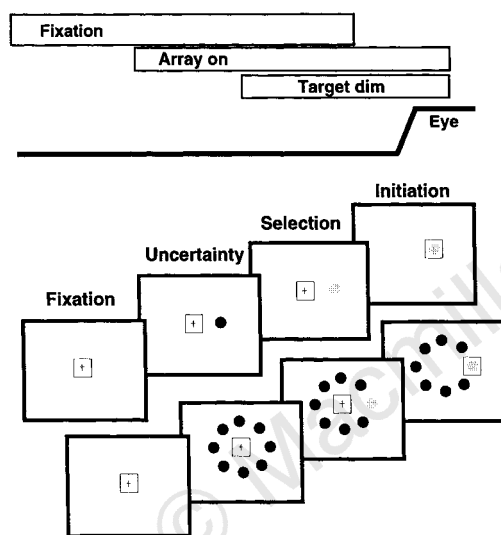


Figure 1 Behavioural task used for separating target uncertainty from target selection. Across the top of the figure a schematic of the temporal arrangement of the task is shown by the labelled time lines and the schematic eye position during the trials is also indicated (Eye). Below is the spatial arrangement of the stimuli on the tangent screen. Only the one possible target and eight possible target trial types are shown. The fixation point is shown as a cross and the stimulus spots are shown as black filled circles. The selected target is indicated schematically by the grey filled circle. The required eye position is indicated by a small square around the fixation point or the selected target.

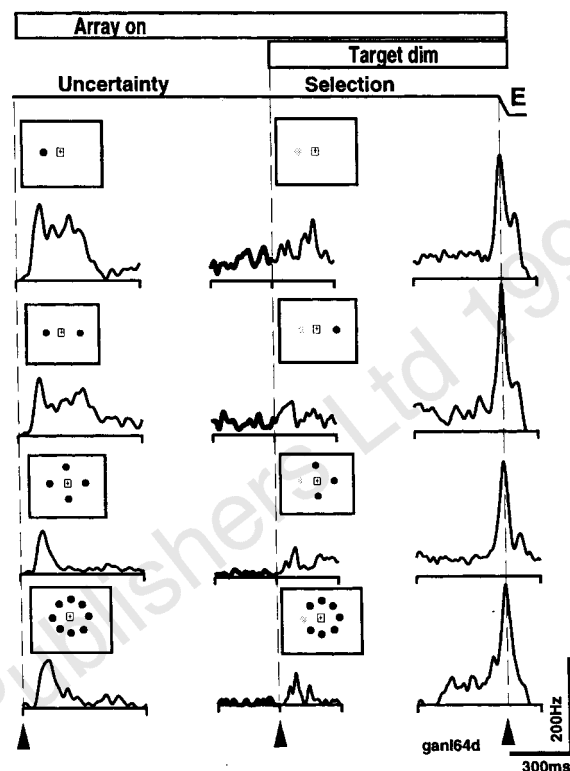
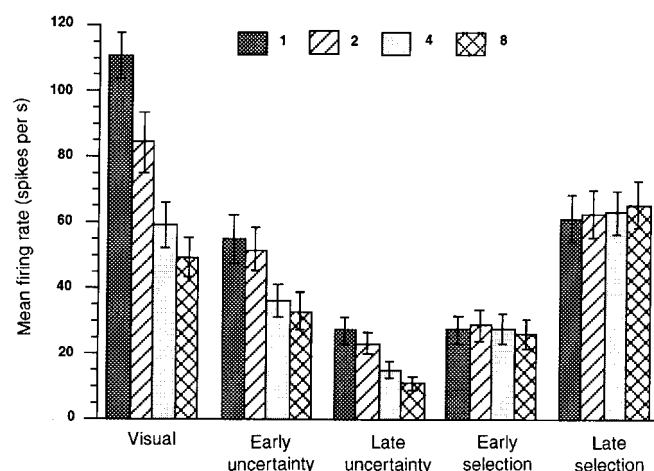


Figure 2 Decreased neuronal activity with increased uncertainty. The activity of a typical buildup neuron studied in the monkey superior colliculus during the three periods of the task is shown by spike density functions averaged over at least 6 trials. The schematic diagrams of the stimulus presentations indicate the trials with one, two, four or eight possible targets. In each column the alignment is indicated by the dashed vertical line and arrowhead. The left column, aligned on array onset, shows the phasic visual on-response and the following lower level of sustained discharge typical of all the neurons studied. The second column shows the shift from uncertainty to selection, and is aligned on target dimming. As the number of possible targets increases, the activity in the period of uncertainty decreases for both the visual on-response and the later sustained response (bold outline) of the buildup neurons. The third column is aligned on saccade initiation; the response amplitude was the same for all trial types. The saccade target was in the centre of the visual field where the burst of the neuron before the saccade was the strongest. The larger visual response with only one possible target (top row, left column) may reflect an enhancement in these saccade-related neurons reminiscent of such enhancement reported for the collicular visual neurons¹⁶. This enhancement would be most prominent with only one possible target because only in this case can the selection (and the accompanying enhancement) be made as soon as the stimulus appears. E is a schematic of eye position.

Figure 3 All neurons showed decreased activity with increased uncertainty. The graph shows the mean activity of 40 neurons from 2 monkeys with one to eight possible targets in the successive periods of the task. Activity decreased as the number of possible targets increased in the visual period (the first 150 ms after the minimal visual latency of intermediate layer SC neurons of 50 ms), in the early period of uncertainty following this visual response (the next 150 ms), and during a period of late uncertainty (200 ms before the target dimmed). In contrast, little change of activity was evident at early target selection (200 ms after the target dimmed) or late target selection (100 ms before the fixation point offset). Bars are 1 standard error of the mean. Note that although some individual neurons showed changes during the selection period (Fig. 2), this was not consistently observed across the sample of neurons (this figure).

All neurons in our sample showed decreasing firing rates during the period of uncertainty as the number of possible targets increased (Fig. 3). The activity decrease from one to eight possible targets was seen in the initial visual period (ANOVA $F(3,156) = 11.81$, $P < 0.001$), the early uncertainty (ANOVA $F(3,156) = 3.43$,

$P < 0.02$), and the late uncertainty period (ANOVA $F(3,156) = 5.69$, $P < 0.001$). This difference was not evident during the selection period (early selection (ANOVA $F(3,156) = 0.06$, $P < 0.98$) or just before the go signal (late selection (ANOVA $F(3,156) = 0.034$, $P < 0.99$)). Thus, the mean response of all 40 neurons in our sample decreased in the period from visual onset to target selection as target uncertainty increased. We also compared the difference between the responses in the one-target and eight-target conditions on a neuron by neuron basis using a Wilcoxon signed rank test. The differences for the intervals before the target selection were also significant on this test (visual, $P < 0.001$; early uncertainty, $P < 0.001$; late uncertainty, $P < 0.001$; early selection, $P < 0.46$; late selection, $P < 0.71$).

In a second experiment, we manipulated the level of target uncertainty by changing the monkey's previous experience on the task. We presented the same eight-stimulus array in two different conditions. In the mixed target condition, any one of the eight possible targets could be selected on any given trial, as in the first experiment. In the blocked trial condition, the same stimulus became the target on every trial. This task had the advantage that in both conditions the stimulus array was identical. Therefore any changes in the activity of neurons could not be attributed to changes in the visual display, but rather to changes in the level of uncertainty the monkeys had about the impending saccadic target. When the target in the response field of the neuron was presented as a saccadic goal repeatedly (blocked target trials) the uncertainty (based on the preceding trials) was minimized and the neuron had a high level of activity (Fig. 4a, top trace and b, top trace). In contrast, when uncertainty was high (mixed target trials) the activity of the neuron diminished (Fig. 4a, bottom trace and b, bottom trace). Quantifying the difference in activity in the blocked and mixed conditions again, using a paired neuron by neuron comparison (Fig. 4c), demonstrated that the modulation of activity in all uncertainty periods was statistically significant (Wilcoxon signed rank test; visual, $P < 0.001$; early uncertainty, $P < 0.001$; late uncertainty, $P < 0.001$) as was the activity during early selection ($P < 0.01$) whereas activity during late selection was not ($P < 0.30$).

Interestingly, the increase in neural activity that occurred in the blocked condition developed over time. For example, after the monkey was switched from mixed to blocked trials, the mean response during the late uncertainty period of the first 10 trials for the neuron depicted in Fig. 4a was 55.05 (s.e.m., 1.57) spikes per s. The mean response increased so that during this interval for the last 10 trials it was 89.65 (s.e.m., 15.70) spikes per s. Across our sample of 32 neurons, the difference in activity during the late uncertainty interval in the first and last 10 trials was statistically significant (Wilcoxon signed rank test, $P < 0.01$).

These experiments show that the activity of many saccade-related superior colliculus neurons changes substantially as the uncertainty of the target changes. Moreover, changes in target uncertainty modulate neuronal activity whether the uncertainty results from changes in the number of possible visual targets or changes in the monkeys' previous experience of which visual stimulus will become the selected target. Thus the buildup of activity that has been used to characterize these neurons' is dependent upon the conditions under which they are being studied. Moreover, the modulation of the vigour of sustained discharge in collicular neurons according to the degree of uncertainty raises the possibility that the change of uncertainty might also influence the activity of neurons in the cerebral cortex that are known to project to the superior colliculus. For example, neurons in the frontal cortex already have been found to alter their activity in the presence of multiple stimuli³. The change in neuronal activity with stimulus selection in other cortical areas not specifically related to the generation of movement, such as those studied in visual attention tasks, might also be altered more or less with changes of uncertainty in the tasks^{11,12}, as seems likely in area V4¹³. Finally, the possibility remains that the activity seen in the superior colliculus before the target selection may represent a

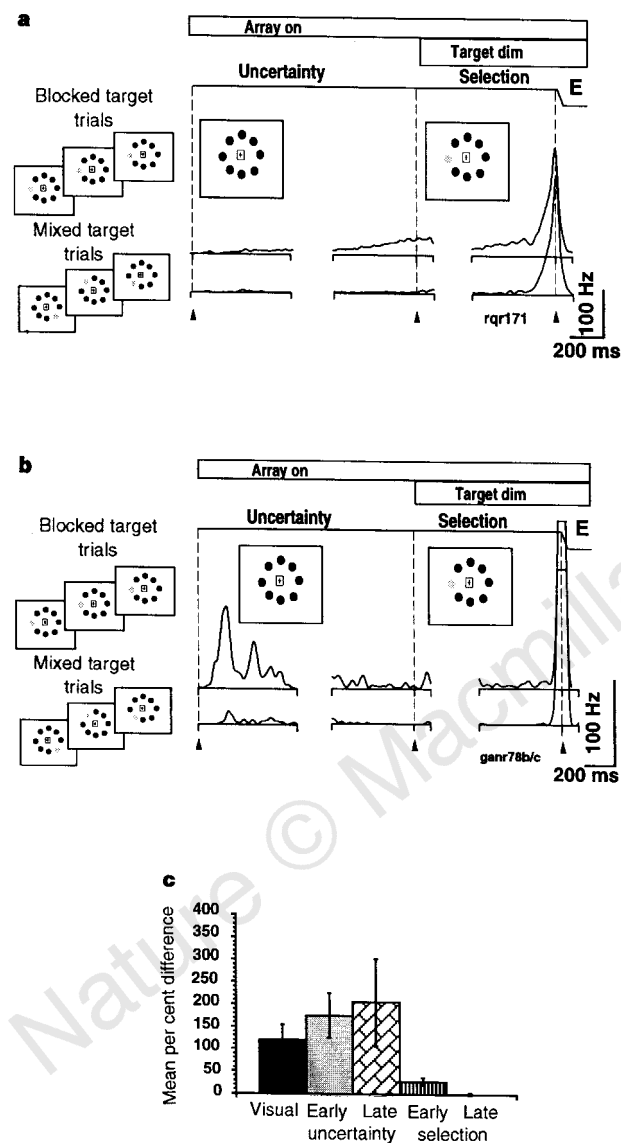


Figure 4 Uncertainty effects persist with individual visual stimuli. **a**, Comparison of the response of a single neuron for trials in which the target selected was always the same one of the eight possible targets (blocked trials) or varied among the eight possible targets (mixed target trials). Each spike density function is an average of 30 trials. The alignment is the same as in Fig. 3. In this task, the selected target and the fixation point offset occurred simultaneously. The neuron in this example had very little visual activity in the eight-target condition. **b**, Same as in **a** except this neuron had a more prominent visual response. Each spike density function is an average of 50 trials. **c**, Mean per cent difference (mean response blocked – mean response mixed/mean response mixed $\times 100$) in activity of 32 neurons from two monkeys in the blocked and mixed target trials. Bars are 1 standard error of the mean. The intervals over which we quantified were the same as in Fig. 3. Because the target dimming and the fixation point turning off occurred simultaneously in this task the two periods of selection largely overlap. Note also that the differences in the neural activity during the different phases of uncertainty were variable across our sample of neurons (compare **a** and **b**), resulting in a large range of difference values. E is a schematic of eye position.

preliminary selection that is then either enhanced or suppressed by descending cortical influences. □

Methods

In two monkeys, we used standard electrophysiological techniques to record single neurons from the superior colliculus, monitor eye movements, and construct spike density histograms as described previously⁷. All experimental protocols were approved by the Institute of Animal Care and Use Committee and complied with Public Health Service policy on the humane care and use of laboratory animals. In all the trials, the monkey followed three steps. First, a central fixation point (projected light-emitting diode spot) came on in the centre of the screen, and the monkey was required to look at it to initiate the trial. Second, the array of possible targets came on. This was the period of uncertainty because which of the spots will become the target was not yet known to the monkeys. One, two, four or eight spots of light were generated by a TV projector, and these trial types were randomly interleaved. One of these spots was always located at the position in the field that yielded the maximal saccade-related activity of the neuron. All other targets were placed equally eccentric but in different directions (in the four cardinal and four oblique directions). Third, one of the spots dimmed. This was the period of selection because at this time the monkeys knew which of the spots was the target for the saccade. The removal of the fixation point served as a 'go' signal and the monkeys were then allowed to make a saccade (initiation). A liquid reward was given if the monkeys maintained their eye position at the target for 300 ms. The selection task was kept simple with a clear target change so that it was essentially a 'popout' task¹⁴ requiring only target detection, not discrimination¹⁵.

In the first experiment (Figs 1–3) the durations of successive periods were randomized among 800–1,200 ms to avoid timing behaviour by the monkey. The periods shown in Fig. 2 are 600 ms so there is no temporal overlap between the successive traces. Because we see the prominent effect of uncertainty in all of these neurons long before saccade onset, it makes little difference whether any given neuron is classified as a prelude burster or a buildup neuron, whether these neuronal categories largely overlap. In the second experiment (Fig. 4) only the probability that the target will be in the movement field of the neuron changes: 100% in blocked trials; 12.5% in mixed target trials. The level of uncertainty is determined by the monkey's experience in previous trials of a consistently located target or a randomly located target.

Received 30 January; accepted 17 June 1997.

1. Maunsell, J. H. R. The brain's visual world: representation of visual targets in cerebral cortex. *Science* **270**, 764–769 (1995).
2. Schall, J. D. & Hanes, D. P. Neural basis of saccade target selection in frontal eye field during visual search. *Nature* **366**, 467–469 (1993).
3. Schall, J. D., Hanes, D. P., Thompson, K. G. & King, D. J. Saccade target selection in frontal eye field of macaque. I. Visual and premovement activation. *J. Neurosci.* **15**, 6905–6918 (1995).
4. Kusunoki, M. & Goldberg, M. E. Responses of parietal visual neurons to attentionally significant stable objects. *Soc. Neurosci. Abstr.* **21**, 665 (1995).
5. Shadlen, M. N. & Newsome, W. T. Motion perception: Seeing and deciding. *Proc. Natl Acad. Sci. USA* **93**, 628–633 (1996).
6. Glimcher, P. W. & Sparks, D. L. Movement selection in advance of action in the superior colliculus. *Nature* **355**, 542–545 (1992).
7. Munoz, D. P. & Wurtz, R. H. Saccade-related activity in monkey superior colliculus. I. Characteristics of burst and buildup cells. *J. Neurophysiol.* **73**, 2313–2333 (1995).
8. Kustov, A. A. & Robinson, D. L. Shared neural control of attentional shifts and eye movements. *Nature* **384**, 74–77 (1996).
9. Sparks, D. L. & Hartwich-Young, R. In *The Neurobiology of Saccadic Eye Movements, Reviews of Oculomotor Research*, Vol. III (eds Wurtz, R. H. & Goldberg, M. E.) 213–256 (Elsevier, Amsterdam, 1989).
10. Miyashita, N. & Hikosaka, O. Minimal synaptic delay in the saccadic output pathway of the superior colliculus studied in awake monkey. *Exp. Brain Res.* **112**, 187–196 (1996).
11. Moran, J. & Desimone, R. Selective attention gates visual processing in extrastriate cortex. *Science* **229**, 782–784 (1985).
12. Treue, S. & Maunsell, J. H. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* **382**, 539–541 (1996).
13. Motter, B. C. Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J. Neurophysiol.* **70**, 909–919 (1993).
14. Treisman, A. M. & Gelade, G. A feature-integration theory of attention. *Cogn. Psych.* **12**, 97–136 (1980).
15. Bravo, M. J. & Makayama, K. The role of attention in different visual-search tasks. *Percept. Psychophys.* **51**, 465–472 (1992).
16. Goldberg, M. E. & Wurtz, R. H. Activity of superior colliculus in behaving monkeys. II. Effect of attention on neuronal responses. *J. Neurophysiol.* **35**, 560–574 (1972).

Acknowledgements. We thank the Laboratory of Diagnostic Radiology Research for magnetic resonance images. We thank J. Steinberg for secretarial assistance preparing the manuscript. We gratefully acknowledge the critical comments made on previous versions by J. Gottlieb and M. A. Sommer.

Correspondence and requests for materials should be addressed to M.A.B. (e-mail: mab@lsr.nei.nih.gov).

PrP-expressing tissue required for transfer of scrapie infectivity from spleen to brain

Thomas Blättler*, Sebastian Brandner*, Alex J. Raeber†, Michael A. Klein*, Till Voigtländer*, Charles Weissmann† & Adriano Aguzzi*

* Institute of Neuropathology, Department of Pathology, University of Zürich, Schmelzbergstrasse 12, CH-8091 Zürich, Switzerland

† Institute of Molecular Biology, University of Zürich, Hönggerberg, CH-8093 Zürich, Switzerland

Much available evidence points to a pathological isoform of the prion protein PrP being the infectious agent that causes transmissible spongiform encephalopathies, but the mechanisms controlling the neurotropism of prions are still unclear. We have previously shown that mice that do not express PrP (*Prnp*^{0/0} mice) are resistant to infection by prions^{1,2}, and that if a *Prnp*^{+/-} neurograft is introduced into such animals and these are infected intracerebrally with scrapie, the graft but not the surrounding tissue shows scrapie pathology³. Here we show that PrP-expressing neurografts in *Prnp*^{0/0} mice do not develop scrapie histopathology after intraperitoneal or intravenous inoculation with scrapie prions. Prion titres were undetectable in spleens of inoculated *Prnp*^{0/0} mice, but were restored to wild-type levels upon reconstitution of the host lymphohaemopoietic system with PrP-expressing cells. Surprisingly, however, i.p. or i.v. inoculation failed to produce scrapie pathology in the neurografts of 27 out of 28 reconstituted animals, in contrast to intracerebral inoculation. We conclude that transfer of infectivity from the spleen to the central nervous system is crucially dependent on the expression of PrP in a tissue compartment that cannot be reconstituted by bone marrow transfer. Thus the requirement for the normal isoform of PrP in peripheral tissues represents a bottleneck for the spread of prions from peripheral sites to the central nervous system.

We placed neurografts derived from wild-type mice or transgenic mice overexpressing PrP (*tga20*; ref. 4) into the brains of *Prnp*^{0/0} mice and delivered scrapie prions (6.5 log LD₅₀ units, where LD₅₀ is the half-maximal lethal dose) by intraperitoneal (i.p.) or intravenous (i.v.) injection. Neither clinical disease nor the histopathological changes characteristic of spongiform encephalopathy developed during the observation period, up to 401 days after inoculation. Histoblots⁵ revealed no protease-resistant PrP^{Sc} in *Prnp*^{0/0} mice carrying *tga20* (*n* = 17) or wild-type (*n* = 2) neurografts. In contrast, the same amount of infectious agent produced scrapie when administered i.p. to 3/3 wild-type mice (incubation time, 209 ± 0 days) or 5/5 *tga20* mice (incubation time, 85 ± 2 days) (Table 1).

These results indicate that the spread of prions from the periphery to the brain of neurografted *Prnp*^{0/0} mice is severely impaired, if not abolished. Three non-exclusive explanations can be offered: (1) prions administered peripherally must first replicate in peripheral organs, from where they invade the central nervous system (CNS); (2) prion spread is possible in the absence of peripheral PrP^C, but does not occur in *Prnp*^{0/0} mice owing to the humoral immune response to PrP that occurs in such animals⁶; (3) whether they replicate or not, prions need a continuous chain of PrP^C-expressing tissues for centripetal propagation within the host.

Possibility (1) predicates that prions administered i.p. must undergo some replication in extracerebral compartments (a process that is impaired in *Prnp*^{0/0} mice⁷) in order to reach the brain. In wild-type mice, the only tissue outside the CNS in which scrapie

prions have been detected is the lymphoreticular system (LRS). It has so far been unclear whether this is due to replication or to accumulation of 'spill-over' infectivity generated in the CNS. LRS colonization by prions is thought to be pathogenetically important: i.p. infection of SCID mice, which have severe LRS deficiencies, is inefficient in generating brain disease⁸⁻¹⁴. Moreover, partial ablation of the LRS through splenectomy increases the incubation time of scrapie through the i.p. route, albeit marginally¹⁵.

To clarify the role of PrP expression in the LRS for neuroinvasion, we lethally irradiated and reconstituted 49 neurografted *Prnp*^{0/0} mice with lymphohaemopoietic stem cells (LSCs) derived from *tga20* (*n* = 17), from wild-type mice (*n* = 21), or from *Prnp*^{0/0} mice

(*n* = 11). Two wild-type mice were reconstituted with *Prnp*^{0/0}-derived LSCs, and developed scrapie 197 and 206 days after i.p. inoculation, possibly owing to residual PrP-expressing spleen cells. Efficiency of reconstitution was assessed *in vivo* by polymerase chain reaction (PCR) and FACS (fluorescence-activated cell sorting) analysis of blood and tail tissue and post mortem by FACS, PCR, and PrP immunohistochemistry on spleen (Fig. 1). Six to eight weeks after transfer, reconstitution ranged between 75% and 100% in peripheral blood. Thirty-six *Prnp*^{0/0} mice underwent adoptive haemopoietic transfer 1-10 weeks before neurografting; five mice were neurografted first and subjected to haemopoietic transfer 1-5 weeks later. Eight to 15 weeks after reconstitution or 5-9 weeks after

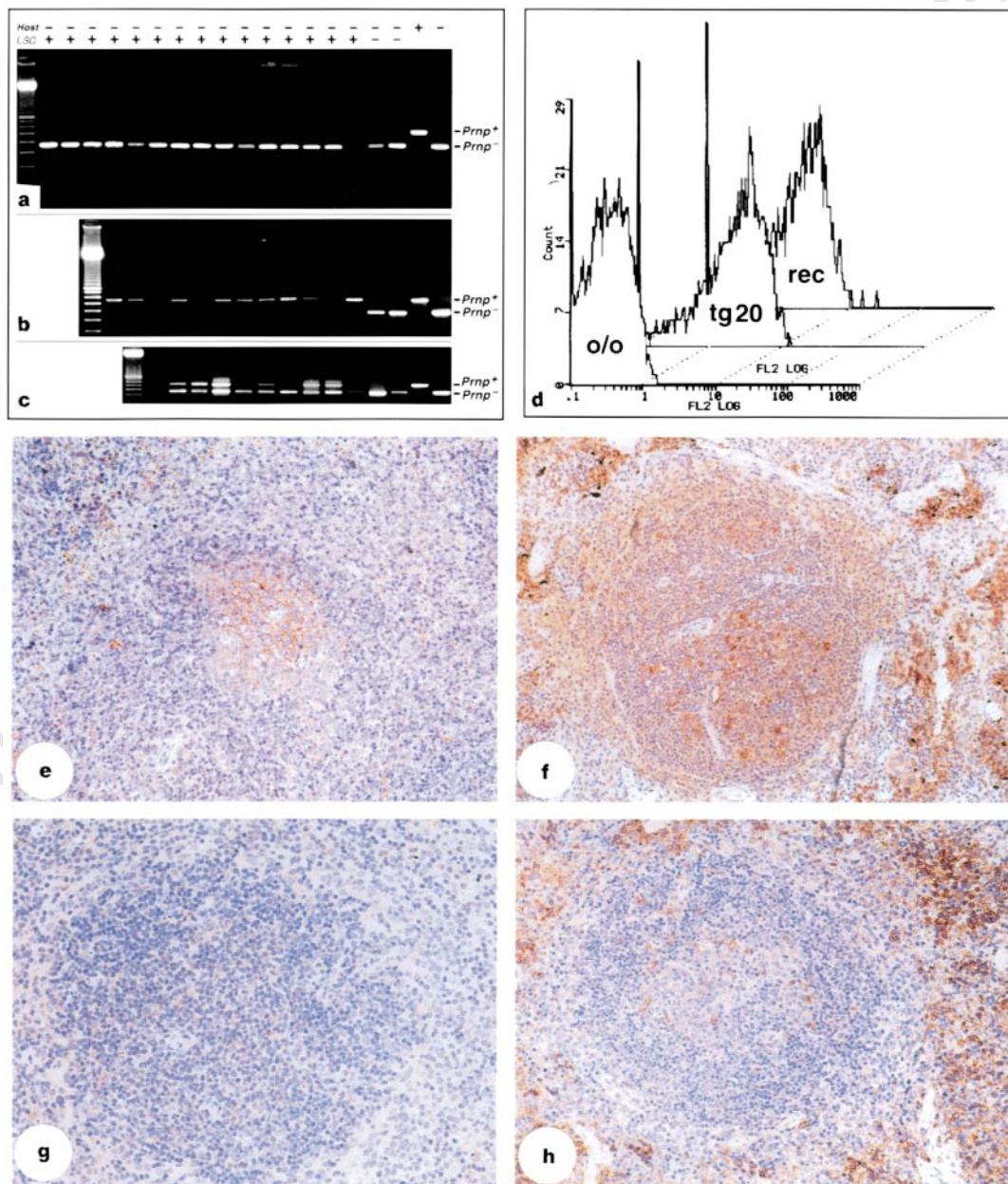


Figure 1 Assessment of the efficacy of reconstitution of the lymphohaemopoietic system by adoptive transfer of LSCs or of T-cell-depleted bone marrow cells. **a-c**, PCR analysis of DNA from peripheral tail tissue (**a**), blood (**b**) and spleen (**c**) of *Prnp*^{0/0} mice reconstituted with wild-type, *tga20* and *Prnp*^{0/0} LSCs. Only the reconstituting genotype can be identified in the peripheral blood of most mice, whereas both bands are identifiable in the spleen, indicating that the stromal component has not been replaced by adoptive transfer. Reconstituting cells produce a faint signal even in tail DNA, probably owing to vertebral haemopoiesis.

Similar results were obtained with wild-type and *tga20* donors. **d**, Assessment of PrP expression on peripheral blood cells by FACS analysis. The signal in *tga20* mice overrides the background noise in *Prnp*^{0/0} mice (*o/o*) by approximately 1 log. The signal in *Prnp*^{0/0} mice reconstituted with *tga20* LSCs (*rec*) is superimposable on that from *tga20* mice. **e-h**, Immunohistochemical assessment of PrP expression in spleen from wild-type mice (**e**), *tga20* mice (**f**), *Prnp*^{0/0} mice (**g**) and *Prnp*^{0/0} mice reconstituted with *tga20* LSCs (**h**).

neurografting (whichever came later), mice were inoculated i.p. with prions (Rocky Mountain Laboratory; 6.5 log LD₅₀) or mock inoculum and maintained for 16–396 days.

To determine prion titres, spleen homogenates from LSC-reconstituted, neurografted ($n = 20$), non-neurografted ($n = 2$) and control mice ($n = 3$) were inoculated intracerebrally into four *tga20* mice per sample. As shown in Table 2, no infectivity was detected in spleens from RML-inoculated *Prnp*^{0/0} mice reconstituted with *Prnp*^{0/0} LSCs ($n = 3$). However, inoculation of neurografted *Prnp*^{0/0} mice reconstituted with wild-type or with *tga20*-derived LSCs led to rapid prion accumulation in spleen, which reached about 7.5 log LD₅₀ units g⁻¹ at 3 weeks after inoculation and remained at about 5.5 log LD₅₀ units or higher at all times thereafter (40 weeks). These titres were similar to those seen in i.p.-infected wild-type mice (Table 2). These results indicate that prion replication occurred in the LRS, did not take place in the absence of PrP-expressing cells in the spleen, and could be fully restored by LSC recolonization with PrP-expressing lymphohaemopoietic stem

cells. Alternatively, because prion titres reach an early plateau and then decrease slightly, transfer of PrP-expressing lymphocytes may have restored a scavenging mechanism of transport to spleen (but not replication) of inoculum administered i.p. The identity of the subset of LSC-derived cells that support prion replication remains unclear.

Histopathological ($n = 15$) and histoblot analysis ($n = 13$) failed to reveal spongiform encephalopathy and/or PrP^{Sc} accumulation in any neurograft of the LSC-reconstituted mice, except for one that had been killed 46 weeks after neurografting and 38 weeks after inoculation i.p. In contrast, after i.c. inoculation, scrapie histopathology was detected in the grafts of all engrafted mice (26 unmanipulated³ and three reconstituted with *tga20* LSCs). Therefore, even the presence of a continuous source of high-titre infectivity in an extremely well irrigated organ such as the spleen (and presumably the entire LRS), although it is perhaps necessary, is certainly not sufficient to mediate prion neuroinvasion.

To assess possibility (2), we searched for antibodies against PrP in

Figure 2 Analysis of immune response to PrP in grafted animals. **a–f**, Western blots containing purified recombinant PrP ('r'), brain homogenate derived from *Prnp*^{0/0} ('o') and *tga20* ('+') mice, respectively, were incubated with sera from *Prnp*^{0/0} mice reconstituted with wild-type (**a**), *tga20* (**b**), or *Prnp*^{0/0} LSCs (**c**), and visualized by enhanced chemiluminescence. Presence of PrP-specific antibodies in the serum is indicated by 27K–30K bands in lane 'r' and by a cluster of bands present in lane '+' but absent from lane 'o'. **d–f**, As positive controls, western blots were stripped and reprobed with polyclonal antibody R340 to recombinant PrP (Y. Kobayashi, unpublished). *Prnp*^{0/0} mice that had received *Prnp*^{0/0} LSCs developed a strong humoral immune response to PrP (**c**), but *Prnp*^{0/0} mice reconstituted with either wild-type (**a**) or *tga20* (**b**) LSCs were largely (**a**) or completely (**b**) tolerant to PrP.

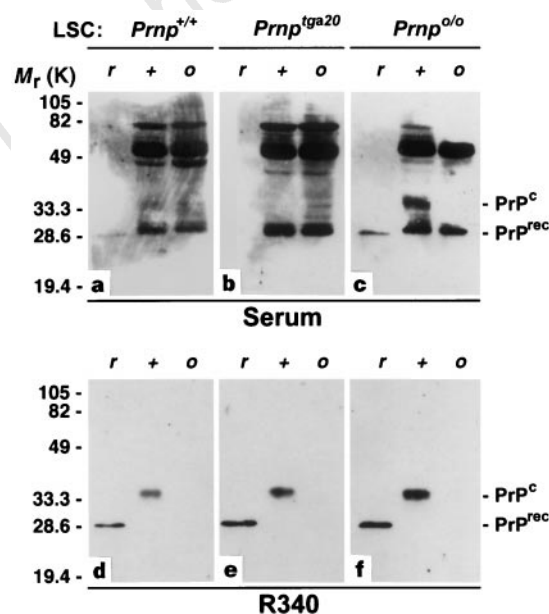


Table 1 Effect of reconstitution with lymphohaemopoietic stem cells on development of spongiform encephalopathy in neurografts 150–396 days after i.p. inoculation of RML scrapie prions

Host genotype	Reconstituting stem cells	Inoculation with prions	Graft spongiform encephalopathy/animals analysed	Time of analysis (days after inoculation)
<i>Prnp</i> ^{0/0}	–	RML i.c.	26/26*	78–498
<i>Prnp</i> ^{0/0}	–	RML i.v.	0/3	242–269
<i>Prnp</i> ^{0/0}	–	RML i.p.	0/14	169–401
<i>Prnp</i> ^{0/0}	–	Mock i.p.	0/3	210–270
<i>Prnp</i> ^{0/0}	<i>Prnp</i> ^{0/0}	RML i.p.	0/8	269–345†
<i>Prnp</i> ^{0/0}	<i>Prnp</i> ^{+/+}	RML i.p.	1/17‡	154–396
<i>Prnp</i> ^{0/0}	<i>Tga20</i>	RML i.c.	3/3	140–250
<i>Prnp</i> ^{0/0}	<i>Tga20</i>	RML i.v.	0/4	201–269
<i>Prnp</i> ^{0/0}	<i>Tga20</i>	RML i.p.	0/7‡	164–331
<i>Prnp</i> ^{0/0}	<i>Tga20</i>	Mock i.p.	0/2	142, 290
			Generalized spongiform encephalopathy/animals analysed	Time to terminal disease (days after inoculation)
<i>Prnp</i> ^{+/+}	–	RML i.p.	3/3	209 ± 0
<i>Prnp</i> ^{+/+}	<i>Prnp</i> ^{0/0}	RML i.p.	2	197, 206
<i>tga20</i>	–	RML i.p.	5/5	85 ± 2
<i>tga20</i>	<i>Tga20</i>	RML i.p.	2/2	92, 112

Grafts were from *Prnp* wild-type mice or from *tga20* mice; there was no difference in the experimental outcome.

* Published data³.

† Some additional mice were killed at earlier time points (Table 2).

‡ Bone marrow reconstitution was performed before neurografting.

Table 2 Scrapie prion titres in spleens of mice after LSC transfer and inoculation with scrapie prions

Host genotype	Bone marrow genotype	Time after inoculation (d)	No. of determinations	Spleen infectivity*
<i>Prnp</i> ^{0/0}	<i>Prnp</i> ^{+/+}	23–39	2	7.3–7.5
		157–203	4	5.4–6.8
		268	2	5.1
		323–390	5	5.4–6.4
		163	4	5.4–6.4
	<i>tga20</i>			
<i>Prnp</i> ^{0/0}	<i>tga20</i> (no brain graft)	168	2	5.5–6.8
<i>Prnp</i> ^{0/0}	<i>Prnp</i> ^{0/0}	17–68	3	Undetectable
<i>Prnp</i> ^{+/+}	–	128	1	6.8
<i>tga20</i>	–	84–89	2	7.5

Mice with the *Prnp* genotype indicated were reconstituted with lymphohaemopoietic stem cells and inoculated i.p. with RML prions. Prion titres in spleen were determined at various times after inoculation.

* Log LD50/g of spleen tissue.

the serum of grafted animals. As before⁶, we found strong immunoreactivity to recombinant *Escherichia coli* PrP and to mouse brain PrP in all engrafted mice (Fig. 2). As this was seen both before and after i.p. inoculation with infectious brain extract, we assume that PrP^C in the graft, rather than PrP^{Sc}, was the offending antigen. Such a vigorous humoral immune response to PrP is unlikely to be 'neutralizing' in the virological sense, inasmuch as it did not prevent graft disease after intracerebral prion inoculation³, despite prolonged breakdown of the blood–brain barrier¹⁶.

It is possible that anti-PrP antibodies prevent the neuroinvasion of prions administered to peripheral sites. To minimize anti-PrP immune responses, we subjected 34 mice to adoptive transfer of wild-type ($n = 21$) or *tga20* ($n = 13$) LSC and subsequent neurografting after 1–10 weeks (24 mice > 4 weeks); in 5 mice, however, the procedure was reversed, with neurografting preceding LSC transfer by 1–5 weeks. Western blot analysis revealed that mice subjected to early adoptive transfer of *tga20* LSC did not develop any detectable immune response to PrP, whereas a brisk immune response was seen when adoptive transfer followed the grafting procedure (data not shown), in agreement with our previous results⁶. As neither group of mice developed neuropathological changes in the graft after i.p. challenge with prions (Table 1), we conclude that the lack of neuroinvasion was not brought about by 'anti-spread' antibodies. However, this experiment does not exclude the possibility that PrP antibodies might exercise an 'anti-spread' activity and so help protect against peripheral inoculation with prions.

According to possibility (3), prions are sessile pathogens that rely on a continuous chain of PrP-expressing tissues to spread from peripheral sites to the CNS⁶. Although in PrP-knockout mice, colonization of the LRS by prions after i.p. inoculation is probably necessary for successful neuroinvasion, it is not sufficient. Neuroinvasion is dependent on PrP expression in a tissue compartment interposed between the LRS and CNS. Indirect evidence suggests that this compartment may comprise parts of the peripheral nervous system^{17,18}. PrP expression in this compartment may constitute a rate-limiting check on prion spread: therefore, strategies aimed at interfering with prion neuroinvasion by reducing its PrP expression may prevent disease after peripheral infection. □

Methods

Bone marrow reconstitution. Mice were irradiated with 900 rad. On the following day, 10⁷ fetal liver cells from E14.5 embryos or 10⁷ T-cell-depleted bone marrow cells from adult mice (genotype: wild-type (C57/B16 × 129SV) or PrP-overexpressing, *tga20*) were administered through the tail vein. Mice were supplied with chloramphenicol in their drinking water, and eight weeks after grafting the extent of reconstitution was assessed by FACS analysis of peripheral blood cells using rabbit antibody R340 to mouse PrP (Y. Kobayashi, unpublished) and a secondary FITC-labelled goat–anti-rabbit antibody. Similar analyses were done on spleen cells post mortem.

In addition, the proportion of reconstituting cells was estimated by competitive PCR on blood and spleen DNA. PCR was carried out with primers P3 (ATT CGC AGC GCA TCG CCT TCT ATC GCC), complementary to the 3' end of the Neo cassette of the knockout allele⁴; P10 (GTA CCC ATA ATC AGT GGA ACA AGC CCA GC), complementary to the 5' end of the *Prnp* reading frame; and 3'NC (ATA ACC CCC TCC CCC AGC CTA GAC CAC GAG), complementary to genomic sequences adjacent to the 3' end of the *Prnp* reading frame. Amplification with primer pair P10–3'NC resulted in a 550-bp DNA fragment; amplification with primer pair P3–3'NC resulted in a 350-bp DNA fragment. Percentage reconstitution was measured against amplified mixtures of *Prnp*^{+/+} and *Prnp*^{0/0} DNA in different proportions using a FluorImager (Molecular Dynamics) and ImageQuant software.

Inoculation. Mice were inoculated with a 1% homogenate of RML-prion-infected brain prepared as described¹⁹; 30 µl was used for intracerebral (i.c.) injection; 100 µl was administered i.p. or i.v.

Detection of anti-PrP antibodies. Brain lysates from *tga20* and *Prnp*^{0/0} mice⁴, as well as recombinant *E. coli* PrP (a gift from R. Glockshuber), were electrophoresed through a 12.5% SDS–polyacrylamide gel and transferred to nitrocellulose. Membranes were incubated with sera from infected animals (diluted 1:200). Visualization was achieved with peroxidase-labelled secondary antiserum and enhanced chemiluminescence (ECL, Amersham).

Analysis of grafts. Tissue was fixed and inactivated with formic acid as described²⁰, embedded in paraffin and subjected to conventional staining and to immunostaining for PrP and for glial fibrillary acidic protein according to standard procedures. Grafts were included in the analysis if their size exceeded 200 × 200 µm and they were detectable on more than 20 adjacent sections. Histoblots were performed as described^{3,5}.

Infectivity assays. Brain and spleen homogenates (10% in 0.32 M sucrose) were prepared from infected animals, and 30 µl (both undiluted or diluted 1:100) was administered i.c. to groups of at least four *tga20* mice for each sample. The incubation time until development of terminal scrapie sickness was determined and infectivity titres were calculated using a regression curve^{3,21}.

Received 3 April; accepted 15 July 1997.

- Büeler, H. R. *et al.* Normal development and behaviour of mice lacking the neuronal cell-surface PrP protein. *Nature* **256**, 577–582 (1992).
- Büeler, H. R. *et al.* Mice devoid of PrP are resistant to scrapie. *Cell* **73**, 1339–1347 (1993).
- Brandner, S. *et al.* Normal host prion protein necessary for scrapie-induced neurotoxicity. *Nature* **379**, 339–343 (1996).
- Fischer, M. *et al.* Prion protein (PrP) with amino-proximal deletions restoring susceptibility of PrP knockout mice to scrapie. *EMBO J.* **15**, 1255–1264 (1996).
- Taraboulos, A. *et al.* Regional mapping of prion proteins in brain. *Proc. Natl Acad. Sci. USA* **89**, 7620–7624 (1992).
- Brandner, S. *et al.* Normal host prion protein (PrP^C) required for scrapie spread within the central nervous system. *Proc. Natl Acad. Sci. USA* **93**, 13148–13151 (1996).
- Sailer, A., Büeler, H., Fischer, M., Aguzzi, A. & Weissmann, C. No propagation of prions in mice devoid of PrP. *Cell* **77**, 967–968 (1994).
- Fraser, H. & Dickinson, A. G. Pathogenesis of scrapie in the mouse: the role of the spleen. *Nature* **226**, 462–463 (1970).
- Kitamoto, T., Muramoto, T., Mohri, S., Doh ura, K. & Tateishi, J. Abnormal isoform of prion protein accumulates in follicular dendritic cells in mice with Creutzfeldt–Jakob disease. *J. Virol.* **65**, 6292–6295 (1991).
- O'Rourke, K. I., Huff, T. P., Leathers, C. W., Robinson, M. M. & Gorham, J. R. SCID mouse spleen does not support scrapie agent replication. *J. Gen. Virol.* **75**, 1511–1514 (1994).
- Lasmezas, C. I. *et al.* Immune system-dependent and -independent replication of the scrapie agent. *J. Virol.* **70**, 1292–1295 (1996).

12. Brown, K. L., Stewart, K., Bruce, M. E. & Fraser, H. in *Transmissible Subacute Spongiform Encephalopathies: Prion Diseases* (eds Court L. & Dodet, B.) 159–166 (Elsevier, Paris, 1996).
13. Kimberlin, R. H. & Walker, C. A. The role of the spleen in the neuroinvasion of scrapie in mice. *Virus Res.* **12**, 201–211 (1989).
14. Fraser, H. & Dickinson, A. G. Studies of the lymphoreticular system in the pathogenesis of scrapie: the role of spleen and thymus. *J. Comp. Pathol.* **88**, 563–573 (1978).
15. Fraser, H. *et al.* Replication of scrapie in spleens of SCID mice follows reconstitution with wild-type mouse bone marrow. *J. Gen. Virol.* **77**, 1935–1940 (1996).
16. Isenmann, S., Brandner, S., Kuhne, G., Boner, J. & Aguzzi, A. Comparative *in vivo* and pathological analysis of the blood–brain barrier in mouse telencephalic transplants. *Neuropathol. Appl. Neurobiol.* **22**, 118–128 (1996).
17. Kimberlin, R. H., Hall, S. M. & Walker, C. A. Pathogenesis of mouse scrapie. Evidence for direct neural spread of infection to the CNS after injection of sciatic nerve. *J. Neurol. Sci.* **61**, 315–325 (1983).
18. Beekes, M., Baldauf, E. & Diringer, H. Sequential appearance and accumulation of pathognomonic markers in the central nervous system of hamsters orally infected with scrapie. *J. Gen. Virol.* **77**, 1925–1934 (1996).
19. Büeler, H. *et al.* High prion and PrP^{Sc} levels but delayed onset of disease in scrapie-inoculated mice heterozygous for a disrupted PrP gene. *Mol. Med.* **1**, 19–30 (1994).
20. Brown, P., Wolff, A. & Gajdusek, D. C. A simple and effective method for inactivating virus infectivity in formalin-fixed tissue samples from patients with Creutzfeldt–Jakob disease. *Neurology* **40**, 887–890 (1990).
21. Prusiner, S. B. *et al.* Measurement of the scrapie agent using an incubation time interval assay. *Ann. Neurol.* **11**, 353–358 (1982).

Acknowledgements. We thank R. Zinkernagel, U. Karrer, E. Flechsig and K. Riem for help with haemopoietic transfer. This study was supported by the Kanton of Zürich, by grants from the Schweizerischer Nationalfonds and Nationales Forschungsprogramm 38 (to A.A. and C.W.) by the European Union, Bundesamt für Veterinärwesen und für Gesundheit and Migros Foundation (A.A.), and by the Human Frontier Science Program (C.W.).

Correspondence and requests for materials should be addressed to A.A. (e-mail: adriano@pathol.unizh.ch).

The mouse *Dazla* gene encodes a cytoplasmic protein essential for gametogenesis

Matteo Ruggiu, Robert Speed, Mary Taggart, Stewart J. McKay, Fiona Kilanowski, Philippa Saunders*, Julia Dorin & Howard J. Cooke

MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

* MRC Reproductive Biology Unit, 37 Chalmers Street, Edinburgh EH3 9EW, UK

RBM and DAZ/SPGY are two families of genes located on the Y chromosome that encode proteins containing RNA-binding motifs, and both have been described as candidate human spermatogenesis genes^{1,2}. Transmission of deletions from father to son has been observed in the case of DAZ^{3–5}, but neither gene family has been shown to be essential for spermatogenesis in human males. The DAZ/SPGY genes are particularly amenable to a knockout approach, as they are found on the Y chromosome in Old World primates and apes⁶, but in other mammals, they are represented only by an autosomal gene, *DAZL*^{7–10}, which is also present in Old World primates and apes. It has also been shown that a *Dazla* homologue is essential for spermatogenesis in *Drosophila*¹¹. Here we show that *Dazla* protein is cytoplasmic in male and female germ cells, unlike the nuclear RBM protein¹². Disruption of the *Dazla* gene leads to loss of germ cells and complete absence of gamete production, demonstrating that *Dazla* is essential for the differentiation of germ cells.

We have expressed the carboxy-terminal 183 amino acids of the mouse *Dazla* in bacteria and raised polyclonal antibodies in rabbits. These antibodies detect an appropriately sized protein on western blots of testis extracts but not on *Dazla*^{Tm1Hgu}/*Dazla*^{Tm1Hgu} extracts in which the gene is disrupted (data not shown). Mouse testis sections stained for *Dazla* show intense staining in germ cells. In the adult testis this is most abundant in pachytene spermatocytes (stages I–VIII). *Dazla* immunostaining was also detectable in type-B spermatogonia, preleptotene and zygotene spermatocytes (Fig. 1a–c), consistent with RNA *in situ* hybridization data¹³. When fluores-

cently labelled antibodies are used, the non-uniform cytoplasmic localization of the protein is evident (Fig. 1d). In the ovary, *Dazla* is present in the maturing follicles (Fig. 1g, h). In embryonic and prepubertal ovary, *Dazla* is cytoplasmic in the oocyte (Fig. 1e, f) and in follicular cells, but in adult follicles the location is consistent with that of the zona pellucida produced by the oocyte^{14,15} or with a peripheral cytoplasmic localization (Fig. 1g, h).

To disrupt the *Dazla* locus we replaced a region of the gene with a neomycin resistance marker (Fig. 2a) by homologous recombination in embryonic stem (ES) cells, which truncates the *Dazla* gene at nucleotide 309, leaving the RNA recognition motif intact. Exons six and seven (which contain the single DAZ repeat) of the gene, and all but five amino acids of exon 5, are deleted, with the next intact exon starting at position 528 in the complementary DNA sequence⁷. ES cells containing this targeted mutation were selected using a polymerase chain reaction (PCR)-based assay and used to generate chimaeric males. These males were mated to MF1 females, and *Dazla*^{Tm1Hgu}/+ mice were intercrossed to give *Dazla*^{Tm1Hgu}/*Dazla*^{Tm1Hgu} animals; Southern blot analysis of DNA from these animals is shown in Fig. 2b. Weaned litter sizes are an average of 10 (9 litters), and we have seen no instances of infertility in *Dazla*^{Tm1Hgu}/+ mice. Transcription of the targeted locus would give rise to a truncated protein, but in a reverse transcription (RT)–PCR assay for the presence of messenger RNA transcribed from the disrupted gene, we were unable to detect transcripts in *Dazla*^{Tm1Hgu}/+ testes where the expressing cell types are present (Fig. 2c). This could be due to message instability or the opposing transcription of the marker gene.

In contrast to *Dazla*^{Tm1Hgu}/+, *Dazla*^{Tm1Hgu}/*Dazla*^{Tm1Hgu} mice are infertile. Four females and three males were test mated at the age of eight or six weeks, respectively (when the *Dazla*^{Tm1Hgu}/+ animals are fertile) with normal males or multiple normal females. No pregnancies were observed, and these experiments were terminated when histological data were obtained. The gross anatomy of the mice appeared to be normal except for the *Dazla*^{Tm1Hgu}/*Dazla*^{Tm1Hgu} animals. Females killed at nine or six weeks after birth had tiny ovaries compared to littermate controls, and histological analysis of these ovaries showed that follicles and ova were absent (Fig. 3c, i). Histological analysis was performed blind to genotype. These females showed evidence of copulatory plugs when caged with wild-type males, and *Dazla*^{Tm1Hgu}/*Dazla*^{Tm1Hgu} males plugged wild-type females. There was a threefold reduction in testis mass in the *Dazla*^{Tm1Hgu}/*Dazla*^{Tm1Hgu} animals compared to wild-type littermates (Table 1), consistent with the loss of germ cells. Epididymal sperm counts were reduced in *Dazla*^{Tm1Hgu}/+ animals, and sperm were not found in *Dazla*^{Tm1Hgu}/*Dazla*^{Tm1Hgu} epididymides (Table 1). Testis sections show severe disruption of testicular histology (Fig. 4), including an almost complete absence of germ cells beyond the spermatogonial stage in *Dazla*^{Tm1Hgu}/*Dazla*^{Tm1Hgu} animals (Fig. 4i–k), and there were fewer tubules per cross-section.

Although heterozygous males showed reduced sperm counts the most striking phenotype is the high level of abnormal sperm produced. We found that 60% of the sperm are visibly abnormal, whereas wild-type littermates show a level of abnormal sperm fourfold lower (Table 1, Fig. 5), but this high level of sperm

Table 1 Testis mass and sperm counts

Genotype	Wild type	<i>Dazla</i> ^{Tm1Hgu} /+	<i>Dazla</i> ^{Tm1Hgu} / <i>Dazla</i> ^{Tm1Hgu}
Mean testis mass (mg)	113	87	33
(s.e.)	(±0.006)	(±0.003)	(±0.004)
No. of testis sampled	8	12	6
Sperm count	65.2	38.6	0
(s.e.)	(±11.3)	(±10.93)	
No. of animals tested	5	7	6
Abnormal sperm (%)	17	61	

Sperm were counted as described¹⁷. Of the abnormal sperm in *Dazla*^{Tm1Hgu}/+ samples, 57% lack heads, 25% have tail abnormalities with small heads, and immature types constitute most of the remainder. Three animals were tested and showed some individual variation.

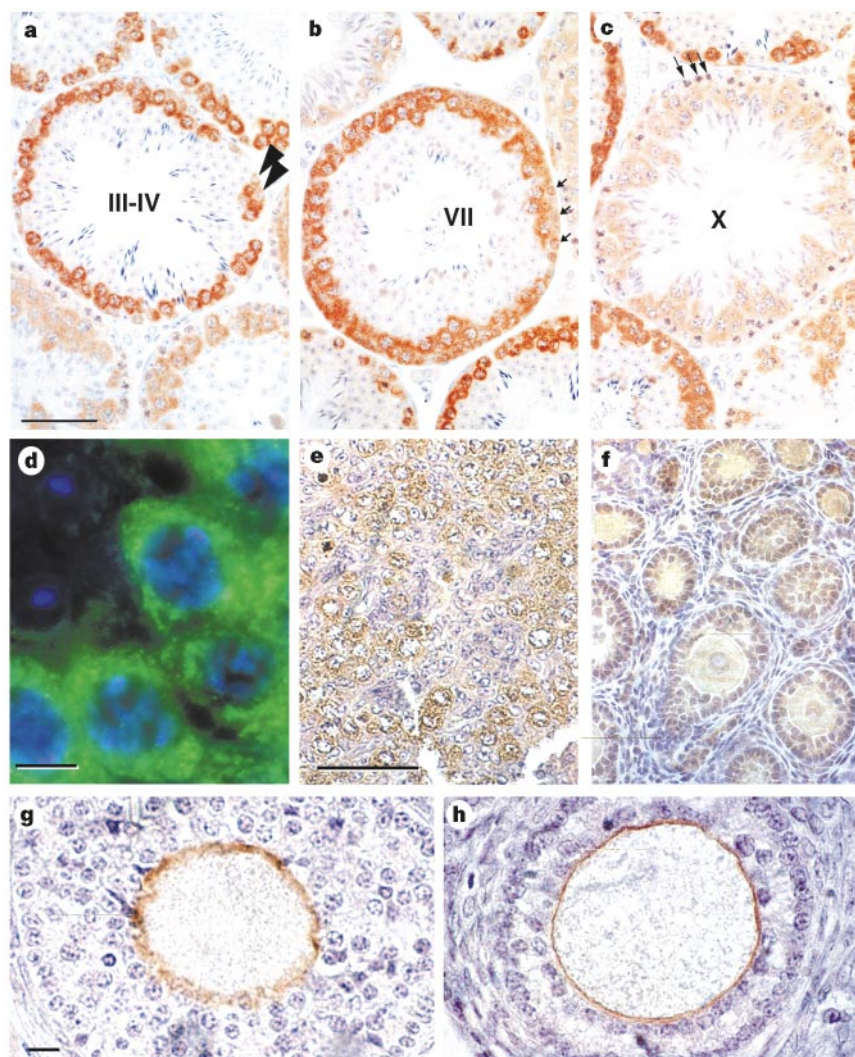


Figure 1 Distribution of Dazla protein in gonads. **a-c**, Adult mouse testis stained with antisera to Dazla (brown) and with haematoxylin (blue). Tubules are in different stages of spermatogenesis. **a**, Stage III-IV with pachytene cells strongly labelled; arrowheads indicate weakly labelled intermediate spermatogonia. **b**, Stage VII arrows indicate preleptotene cells. **c**, Stage X arrows indicate labelled leptotene spermatocytes. **d**, Spermatocytes stained with fluorescently labelled antibody (green). **e**, Fetal ovary at 17 d.p.c. showing intense staining of pachytene cells. **f**, Prepubertal ovary (14 d.p.c.) stained as in **a**. Cytoplasmic staining is visible in oocyte and follicular cells. **g, h**, Adult ovary follicles stained as in **a**. Staining is cytoplasmic with a peripheral concentration. All staining was completed by the immunizing peptide.

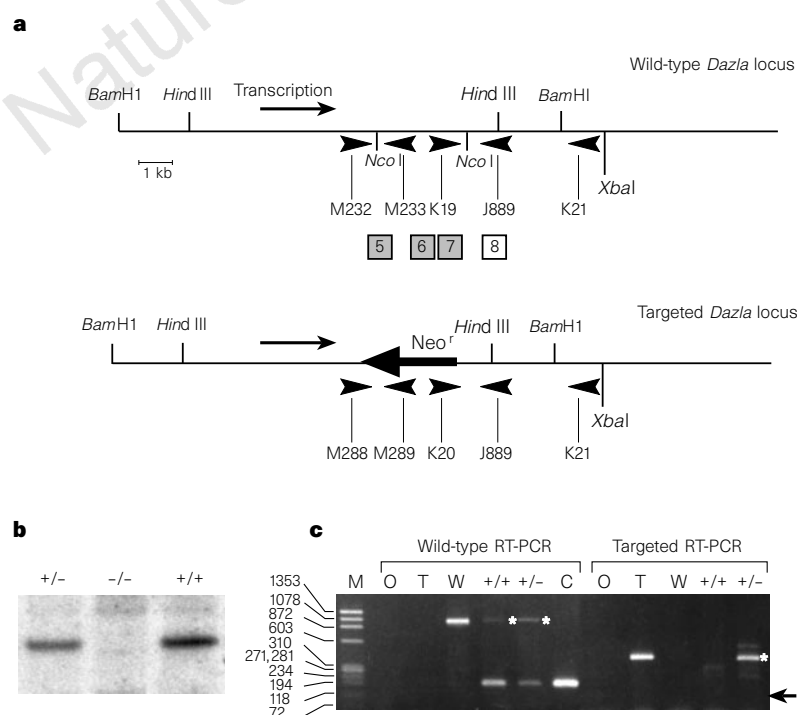


Figure 2. Targeting of *Dazla*. **a**, Wild-type and targeted *Dazla* loci. Transcription directions are indicated by arrows and exons by numbered boxes. PCR primers used are indicated by arrowheads. The targeting construct consists of the *Bam*HI fragment from the wild-type locus with the *Nco*I fragment replaced with a neomycin resistance marker. Exons affected or deleted by this are shaded. **b**, Southern blot analysis of DNA from wild-type (+/+), heterozygous *Dazla*^{Tm1Hgu} (+/-), and homozygous animals (-/-) probed with the *Nco*I fragment deleted from the targeting construct. **c**, Analysis of gene expression of RT-PCR products, showing DNA (O), targeting construct (T) and wild-type (W) DNAs as amplification and DNA contamination controls, a cloned *Dazla* cDNA template (C) and +/- and +/- cDNA from testes of wild-type and heterozygous animals. Asterisks indicate genomic DNA products in cDNA tracks; arrow, the predicted position of the disrupted gene cDNA transcript.

abnormality would not be expected to lead to infertility. This heterozygous phenotype would be consistent with our having created a null mutation if normal levels of the protein were required in spermatogenesis. An equivalent heterozygous phenotype in females would be difficult to detect.

These observations suggest that germ cells are reduced in the gonads, as a result of either reduced proliferation in the fetus or loss of germ cells after proliferation. To distinguish between these possibilities we analysed one litter at 15 days post-coitum (d.p.c.) and one at 19 d.p.c. The *Dazla*^{Tm1Hgu}/*Dazla*^{Tm1Hgu} ovaries at 15 d.p.c. appear relatively normal, with many early pachytene cells (Fig. 3d–f), but by 19 d.p.c. there is a marked deficiency of oocytes, with many of those remaining undergoing atresia (Fig. 3g, h). Testes of 15 d.p.c. *Dazla*^{Tm1Hgu}/*Dazla*^{Tm1Hgu} mice are similar to those of *Dazla*^{Tm1Hgu}/+ animals (Fig. 4a, b), but by 19 d.p.c. they have sparse tubule populations (Fig. 4c–e). Sertoli-cell:germ-cell ratios are increased from 4:1 in wild-type to 8:1 in both *Dazla*^{Tm1Hgu}/+ and *Dazla*^{Tm1Hgu}/*Dazla*^{Tm1Hgu} animals. Meiosis has not initiated by this stage, and so this loss of germ cells cannot be dependent on pachytene arrest.

Our results show that the mouse *Dazla* gene is essential for

development and survival of germ cells in both ovary and testis. Male *Dazla*^{Tm1Hgu}/+ animals produce reduced numbers of sperm with more abnormalities, suggesting a quantitative requirement for Dazla protein. The related *boule* gene in *Drosophila* is essential for spermatogenesis but does not seem to be required for female gametogenesis¹¹, in contrast with the murine *Dazla* gene. In the mouse the effects of the absence of Dazla are seen when germ-cell mitotic divisions cease in both sexes¹⁶. Female germ cells enter meiosis and arrest, and male mitotic division of germ cells is largely suspended until the onset of spermatogenesis. In contrast to the nuclear RBM protein, Dazla is cytoplasmic and so is unlikely to be a transcription or splicing factor, but may be involved in translational control by packaging or localizing mRNAs. The adult location of the protein in germ-cell types that are not present at the time of the effects of gene disruption suggests that Dazla has different functions in embryonic and adult gonads. Human semen normally contains high levels of abnormal sperm comparable to the levels seen in *Dazla*^{Tm1Hgu}/+ males. This could obscure the heterozygous effects of *DAZL* mutations, especially in the presence of DAZ protein. In humans, therefore, heterozygous phenotypes may only be evident in females.

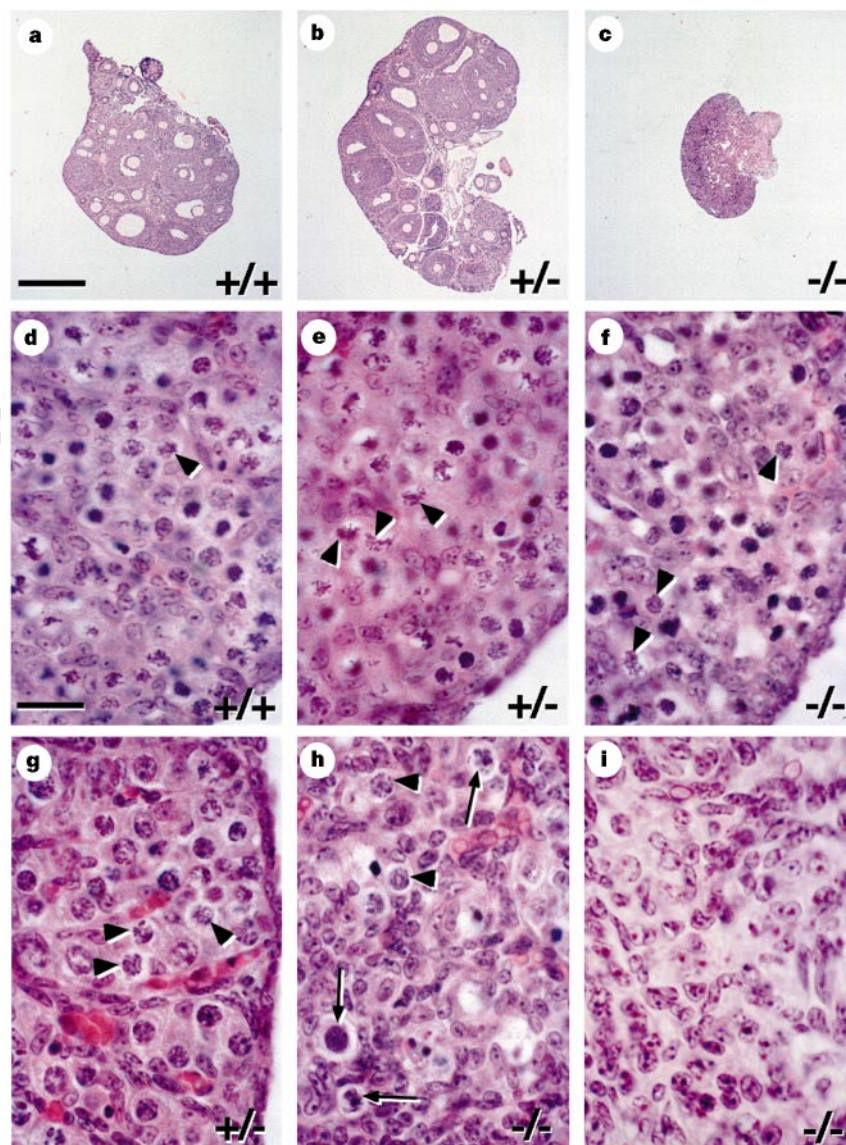
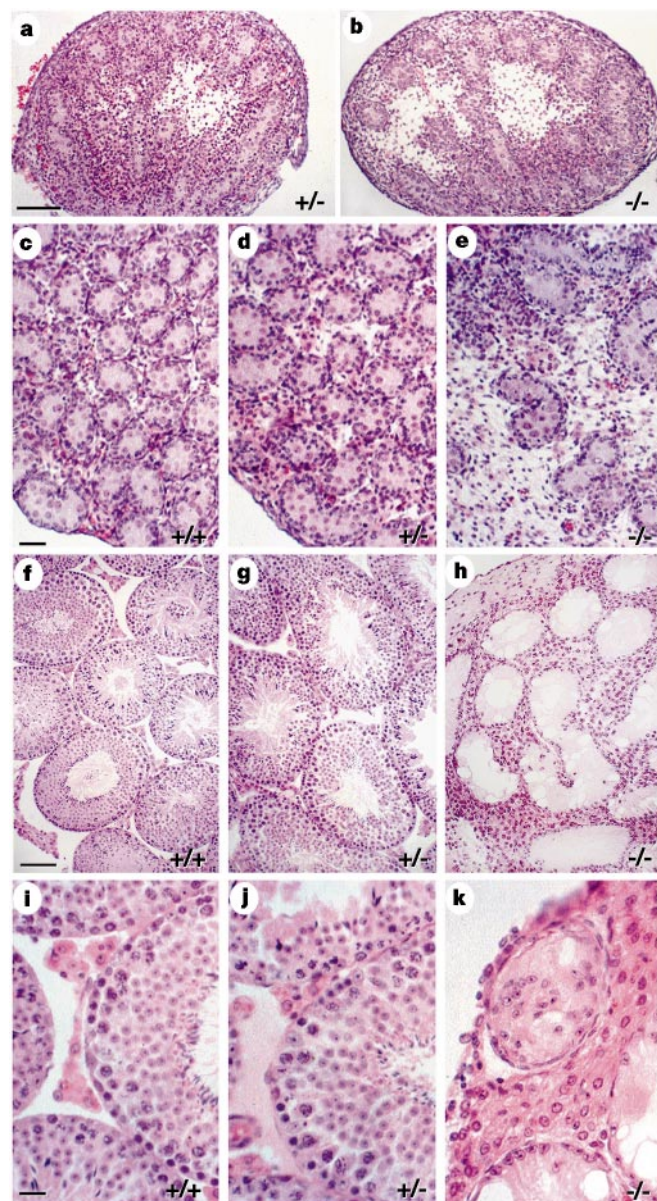


Figure 3 Histological analyses of ovaries. **a–c**, Ovaries from wild-type (+/+), *Dazla*^{Tm1Hgu}/+ (+/-) and *Dazla*^{Tm1Hgu}/*Dazla*^{Tm1Hgu} (-/-) adult (9-week-old) mice. **d–f**, Fetal ovary at 15 d.p.c. with pachytene cells indicated by arrowheads. **g, h**,

Ovaries at 19 d.p.c. showing pachytene cells (arrowheads) or atretic pachytene cells (arrows). **i**, Adult ovary. Scale bars, 400 μ m in **a–c**, and 20 μ m in **d–i**.



Note added in proof: The nomenclature of the autosomal DAZ homologues is in the process of rationalization. □

Methods

Antibody production. The DNA encoding the C-terminal 183 amino acids of Dazla was cloned into pRSET (Invitrogen), and the induced protein was used to raise antisera in rabbits, according to standard methods. Antisera were affinity purified with the bacterially expressed protein.

Targeting and expression. A *Bam*HI fragment was used for targeting in which an *Nco*I fragment of 2.7 kb was replaced by a neomycin resistance marker from pMC1NeoPolyA as a *Xho*I–*Sall* fragment after deletion of the *Bam*HI site in the poly linker. The insert from this plasmid was electroporated into C129/CGR8 ES cells and individual neomycin-resistant clones were picked and expanded for analysis. We developed a PCR assay for targeting using primers k20 and k21 (k20, GCTTCCTCTTGCAAAACCAC; k21, GCACCCAGAGATTACTGATATGG). Two cell lines were injected into blastocysts, and both gave rise to chimaeric transmitting males. The progeny of these males had identical phenotypes. Progeny were typed using primers j889 (CAGTGGCTTTTGAAATTATCA) with either k19 (TCAGCATGCATTCTTTCAGTCTTTC) or k20. PCR to detect transcripts from targeted and non-targeted loci was performed using primers m289/M288 (m288, ATGACGTGGATGTGCAGAAAGAT; m289, GGGATCGCAATAAAAAGACAG) or m232/m233 (m232, GATAATCACTGATCGAAGTGGTGTG; m233, TGGTGGAGGAGGAGGAT-TAAAA), respectively.

Histology. Mice were killed and ovaries and testes fixed in Bouins solution for histological analysis, and tissue sections 6 μm thick were prepared and stained with haematoxylin and eosin by standard methods. Sperm preparations were analysed as described¹⁷. Ages of embryos are given with day of plug as 1 d.p.c.

Received 8 May; accepted 18 July 1997

1. Reijo, R. *et al.* Diverse spermatogenic defects in humans caused by Y-chromosome deletions encompassing a novel RNA-binding protein gene. *Nature Genet.* **10**, 383–393 (1995).
2. Ma, K. *et al.* A Y-chromosome gene family with RNA-binding protein homology candidates for the azoospermia factor AZF controlling human spermatogenesis. *Cell* **75**, 1287–1295 (1993).
3. Reijo, R., Alagappan, R. K., Patrizio, P. & Page, D. C. Severe oligozoospermia resulting from deletions of azoospermia factor gene on Y-chromosome. *Lancet* **347**, 1290–1293 (1996).

◀ **Figure 4** Testis histology. Genotypes are as described in the legend to Fig. 3. **a, b**, Fetal testes at 15 d.p.c. showing no clear differences between mutant and heterozygous animals. **c–e**, Testes at 19 d.p.c. showing clear differences between heterozygous and homozygous mutant animals. The sparsity of tubules in **e** is reflected in serial sections throughout the testis (not shown). **f–h**, Testes at 9 weeks *post partum* showing absence of germ cells in mutant animals. **i–k**, Tubules from **f–h** at greater magnification. Scale bars, 100 μm in **a–e**, 500 μm in **f–h**, and 50 μm in **i–k**.

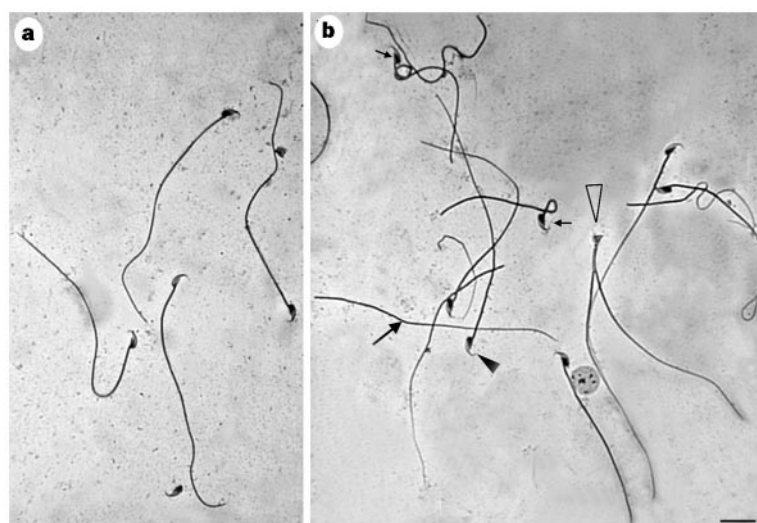


Figure 5 Abnormal sperm production in heterozygotes. **a**, Wild-type sperm. **b**, *Dazla*^{Tm1Hgu/+} sperm. Normal (filled arrowhead), abnormal head with two tails (open arrowhead), abnormal and truncated tails (small arrow) and tail only (large arrow). Scale bar, 100 μm.

4. Vogt, P. H. *et al.* Human Y-chromosome azoospermia factors (azf) mapped to different subregions in Yq11. *Hum. Mol. Genet.* **5**, 933–943 (1996).
5. Nakahori, Y. *et al.* The Y-chromosome region is essential for spermatogenesis. *Horm. Res.* **46**, 20–23 (1996).
6. Shan, Z. H. *et al.* A SPGY copy homologous to the mouse gene *Dazl* and the *Drosophila* gene *boule* is autosomal and expressed only in the human male gonad. *Hum. Mol. Genet.* **5**, 2005–2011 (1996).
7. Cooke, H. J., Lee, M., Kerr, S. & Ruggiu, M. A. A murine homolog of the human *DAZ* gene is autosomal and expressed only in male and female gonads. *Hum. Mol. Genet.* **5**, 513–516 (1996).
8. Reijo, R. *et al.* Mouse autosomal homolog of *DAZ*, a candidate male-sterility gene in humans, and is expressed in male germ-cells before and after puberty. *Genomics* **35**, 346–352 (1996).
9. Saxena, R. *et al.* The *DAZ* gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. *Nature Genet.* **14**, 292–298 (1996).
10. Yen, P. H., Chai, N. N. & Salido, E. C. The human autosomal gene *dazl* – testis specificity and a candidate for male-infertility. *Hum. Mol. Genet.* **5**, 2013–2017 (1996).
11. Eberhart, C. G., Maines, J. Z. & Wasserman, S. A. Meiotic cell cycle requirement for a fly homologue of human *Deleted in Azoospermia*. *Nature* **381**, 783–785 (1996).
12. Elliott, D. J. *et al.* Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm. *Proc. Natl Acad. Sci. USA* **94**, 3848–3853 (1997).
13. Niederberger, C., Agulnik, A. I., Cho, Y., Lamb, D. & Bishop, C. E. In situ hybridization shows that *Dazl* expression in mouse testis is restricted to premeiotic stages iv–vi of spermatogenesis. *Mamm. Genome* **8**, 227–278 (1997).
14. Bleil, J. D. & Wassarman, P. M. Synthesis of zona pellucida proteins by denuded and follicle-enclosed mouse oocytes during culture in vitro. *Proc. Natl Acad. Sci. USA* **77**, 1029–1033 (1980).
15. Bleil, J. D. & Wassarman, P. M. Structure and function of the zona pellucida: identification and characterization of the proteins of the mouse oocyte's zona pellucida. *Dev. Biol.* **76**, 185–202 (1980).
16. Pelliniemi, L. J., Frojndman, K. & Paranko, J. In *The Sertoli Cell* (eds Russell, L. D. & Griswold, M. D.) 87–114 (Casper River, Clearwater, FL, 1993).
17. Searle, A. G. & Beechey, C. V. Sperm-count, egg-fertilisation and dominant lethality after X-irradiation of mice. *Mut. Res.* **22**, 63–77 (1974).

Acknowledgements. We thank all our colleagues in the Human Genetics Unit who helped in the discussion and production of this paper. This work was supported by the MRC. M.R. is supported by an Italian Telethon grant.

Correspondence and requests for materials should be addressed to H.J.C. (e-mail: howard@hgu.mrc.ac.uk).

Mechanism of inhibition of the human matrix metalloproteinase stromelysin-1 by TIMP-1

Franz-Xaver Gomis-Rüth[†], Klaus Maskos[†], Michael Betz[†], Andreas Bergner[†], Robert Huber[†], Ko Suzuki[‡], Naoki Yoshida[‡], Hideaki Nagase[‡], Keith Brew[§], Gleb P. Bourenkov^{||}, Hans Bartunik^{||} & Wolfram Bode^{*}

^{*} Max-Planck-Institut für Biochemie, Abteilung für Strukturforschung, D-82152 Martinsried, Germany

[‡] Department of Biochemistry and Molecular Biology, University of Kansas Medical Center, Kansas City, Kansas 66160, USA

[§] Department of Biochemistry and Molecular Biology, University of Miami, School of Medicine, Miami, Florida 33101, USA

^{||} AG Proteindynamik MPG-ASMB, c/o DESY, D-22603 Hamburg, Germany

[†] These authors contributed equally to this work.

Matrix metalloproteinases (MMPs) are zinc endopeptidases that are required for the degradation of extracellular matrix components during normal embryo development, morphogenesis and tissue remodelling¹. Their proteolytic activities are precisely regulated by endogenous tissue inhibitors of metalloproteinases (TIMPs)^{1–5}. Disruption of this balance results in diseases such as arthritis, atherosclerosis, tumour growth and metastasis^{1,2}. Here we report the crystal structure of an MMP-TIMP complex formed between the catalytic domain of human stromelysin-1 (MMP-3) and human TIMP-1. TIMP-1, a 184-residue protein⁵, has the shape of an elongated, contiguous wedge. With its long edge, consisting of five different chain regions, it occupies the entire length of the active-site cleft of MMP-3. The central disulphide-linked segments Cys 1–Thr 2–Cys 3–Val 4 and Ser 68–Val 69 bind to either side of the catalytic zinc. Cys 1 bidentally coordinates this zinc, and the Thr-2 side chain extends into the large specificity pocket of MMP-3. This unusual architecture of the interface

between MMP-3 and TIMP-1 suggests new possibilities for designing TIMP variants and synthetic MMP inhibitors with potential therapeutic applications.

The TIMP family at present comprises four members (TIMP-1 to TIMP-4) of relative molecular mass (M_r) ranging from 22K to 30K, with 40–50% sequence identity³. All TIMPs inhibit active MMPs with relatively low selectivity, forming tight non-covalent 1:1 complexes⁴. Native TIMP-1 contains two glycosylation sites⁵. A C-terminally truncated form of TIMP-1 comprising only the N-terminal two-thirds of its polypeptide chain (N-TIMP-1) has been shown to inhibit MMP-3 with a slightly reduced affinity compared with full-length TIMP-1 (refs 4, 6, 7). These experiments suggested that TIMPs interact with MMPs predominantly through their N-terminal moiety. A preliminary nuclear magnetic resonance (NMR) model of N-TIMP-2 has been reported⁸, but the mechanism of MMP inhibition by the TIMPs is not understood. Most MMPs, including stromelysin-1, are secreted as multimodular proenzymes that can be activated, with a 165-residue catalytic domain flanked by an N-terminal activation peptide and a spacer-linked C-terminal domain, respectively. Three-dimensional structures have been determined for a few catalytic MMP domains, including those of stromelysin-1 (MMP-3)^{9–11} and its proform¹⁰, and for full-length porcine MMP-1 (ref. 12). To understand the mechanism of interaction between MMPs and TIMPs, and to allow the design of more specific TIMP species, we have undertaken an X-ray crystal-structure analysis of the complex formed between unglycosylated human TIMP-1 and the catalytic domain of MMP-3.

The X-ray structure of this complex reveals that the wedge-shaped inhibitor slots with its edge into the active-site cleft of the MMP-3 catalytic domain (Fig. 1). The TIMP polypeptide chain folds into a contiguous elongated domain, with the N- and the C-terminal halves forming two opposing 'subdomains' (Fig. 2). As previously found for N-TIMP-2 (ref. 8), the N-terminal subdomain exhibits the oligosaccharide/oligonucleotide binding fold of proteins such as staphylococcal nuclease, yeast aspartyl-tRNA synthetase, some *Escherichia coli* enterotoxins, and a *Bacillus subtilis* cold-shock protein¹³. The detailed geometry of our full-length TIMP-1 molecule, however, deviates significantly from that of the N-TIMP-2 model (whose coordinates were kindly supplied by R. A. Williamson), probably because this early NMR model was still unrefined. The oligosaccharide/oligonucleotide binding-fold region of TIMP-1 (for nomenclature, see Fig. 2) consists of a five-stranded β -pleated sheet rolled into a β -barrel of conical shape. The narrower opening of this barrel is bounded by the flexible B–C loop, and its wider exit is, in contrast to other such proteins, covered by an extended segment, connecting strands C and D, which we therefore designate as a 'connector'.

The N-terminal TIMP-1 segment, tightly fixed to the C-connector loop and to the E–F loop by disulphide bridges Cys 1–Cys 70 and Cys 3–Cys 99 (Figs 1, 2 and 3), forms part of the molecular edge. At Val 4 it turns towards the centre of the molecule, where it passes α -helix 1 before entering the barrel (outlined in Fig. 2) at strand A. After strand F (which in the β -sheet represents a continuation of the fifth strand E; Fig. 2), the polypeptide chain runs through α -helix 2 and 3_{10} -helix 3 before it crosses over to the C-terminal subdomain.

This C-terminal subdomain of TIMP-1 is organized less regularly (Fig. 2). It includes two adjacent β -sheets, made by two parallel (G and H) and two antiparallel β -strands (I and J), and the short α -helix 4, and possesses a minor internal core mainly composed of the two disulphide bridges Cys 127–Cys 174 and Cys 145–Cys 166. The last three TIMP-1 residues are not defined, and presumably form a flexible tail at the TIMP surface. Besides the covalent polypeptide chain, the two subdomains of TIMP-1 are held together through side-chain interactions (Fig. 2). In particular, the strictly conserved residues His 7 and Gln 9 of the N-terminal subdomain pass over to the C-terminal subdomain. Thus the 'surface' of the N-terminal subdomain around His 7, previously suggested to represent a

potential docking site towards the MMPs⁸, forms part of the intramolecular subdomain interface of the TIMPs.

In the complex, six sequentially separate polypeptide segments of TIMP-1 (four provided by the N-terminal and two by the C-terminal subdomain) interact with MMP-3 (Figs 1 and 2). Altogether, surfaces of about 1,300 Å² in each molecule are removed from contact with bulk water upon complex formation. The N-terminal segment and the C-connector loop of TIMP-1 covalently linked by disulphide Cys 1–Cys 70, take part in three-quarters of all intermolecular contacts. The N-terminal segment Cys 1–Thr 2–Cys 3–Val 4 binds to the active-site cleft (subsites S1 to S3') of MMP-3 in a manner very similar to that of the residue before (P1) and the three residues (P1', P2' and P3') after the scissile peptide bond of a bound peptide substrate¹⁴ (Figs 3 and 4). It inserts between two antiparallel MMP-3 strands (peptide backbones of S162–S166 and S220–S223; stromelysin residues assigned with the prefix S) bordering the MMP-3 active-site cleft, and makes further intermolecular contacts with residues at the bottom of this cleft (S198–S202, S205 and S211).

Cys 1 is located directly above the catalytic zinc, with its N-terminal α-amino and its carbonyl group placed at a distance of about 2.0 Å. As well as binding to the zinc, the α-amino group (which is presumably unprotonated) could form hydrogen bonds to

a neighbouring carbonyl oxygen (Ser 68) and to one of the carboxylate oxygens of the 'catalytic' Glu S202 residue of MMP-3 (Fig. 4). The side chain of this Glu S202 moves closer when TIMP binds. Thus the zinc-bound water molecule normally found in non-ligated zinc endopeptidases close to this 'catalytic' Glu is excluded in the complex. This α-amino acyl group superimposes partly with the hydroxamic acid moiety commonly used as a zinc-chelating component in many synthetic MMP inhibitors.

The Thr 2 side chain of TIMP-1 extends into the large S1' specificity pocket of MMP-3 (see Fig. 3); its β-methyl group fits into a hydrophobic niche, and its β-hydroxy group is directed towards the more polar entrance neck of this pocket. The cystine bridge Cys 3–Cys 99 points away from the TIMP edge towards the underlying E–F loop, and the side chains of Val 4 and Pro 5 just touch the MMP. The high sequence conservation of the N-terminal segments of all inhibitory TIMPs predicts this interaction to be common to all MMP–TIMP complexes.

The adjacent C-connector loop of TIMP-1 (Ala 65–Cys 70; see Fig. 2) occupies that part of the active site which in productive MMP–substrate complexes interacts with substrate residues located two and three positions N-terminal of the scissile peptide bond (P2 and P3, to the left of the catalytic zinc; Fig. 3)¹⁴. However, Ser 68 and Val 69 of TIMP-1 are arranged in a nearly opposite orientation to



Figure 1 Stereo ribbon diagram of the complex formed between TIMP-1 (red) and the MMP-3 catalytic domain (blue). Yellow bonds indicate disulphide bridges, and the spheres represent the MMP-3 bound 2 zinc (pink) and 3 calcium (yellow) ions. The active-site cleft of MMP-3 is directed halfway towards the bottom and the viewer. The MMP-3 polypeptide chain starts with a 'cap-like' structure on the left. Cys 1 of TIMP-1 is located on top of the catalytic zinc (centre of the complex). Its

first four residues insert between the 'edge' strand (above, front) and the strand blocking off the specificity pocket (below, back) of MMP-3 (ref. 14), before the TIMP-1 chain deviates from the active-site cleft to form the N-terminal subdomain (left, bottom) and the C-terminal subdomain (right, bottom). Figure drawn with MOLSCRIPT and rendered with RASTER3D.

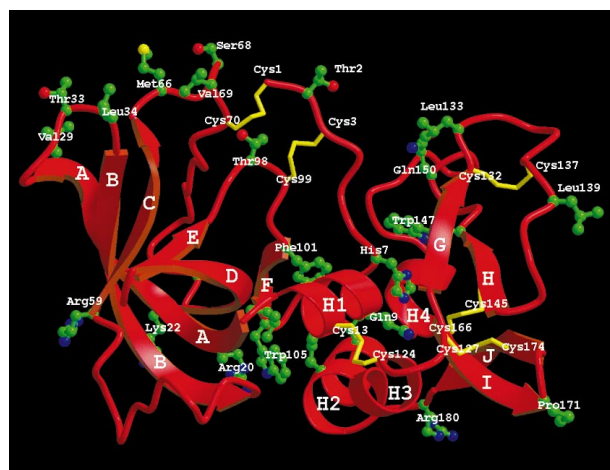


Figure 2 Ribbon of TIMP-1 (red) shown together with the full side chains of some important residues and the disulphide connections. The inhibitor is slightly turned around a horizontal compared with Fig. 1 to allow a view on the 'front' side. The A–B loop, the C-connector loop, the H–I loop and the G–H loop flank the N-terminal segment (top), together forming the molecular 'edge'. The polypeptide chain first forms the N-terminal subdomain (left) starting with Cys 1–Thr 2–Cys 3 (top, centre), turning towards the molecular centre, running through helix H1, forming the five-stranded β-barrel (strands A–F), and ending with helices H2 and H3. The chain then enters the C-terminal subdomain (right), forms the two-stranded G–H β-sheet, runs through a multiple coil section and helix H4, before it ends up in the I–J sheet (right, bottom). Figure drawn with MOLSCRIPT and rendered with RASTER3D.

the P3–P2 segment of a bound peptide substrate. The interactions in this region (made with MMP-3 segments S83–S86, S154–S155, S163–S167, S205 and S210–S211) are predominantly hydrophobic. The side chains of Met 66 and Val 69 in particular extend into a hydrophobic ‘cap-like’ structure formed by the N-terminal segment Phe S83–Pro S90 of MMP-3 (Fig. 1). This ‘cap’ results from the disruption of the salt bridge formed between the Phe S83 α -ammonium group and the Asp S237 carboxyl group, and from the complete rearrangement and up to 15-Å shift of the N-terminal segment of MMP-3, and is accompanied by a large movement of the Phe S210 benzyl side chain away from the MMP-3 substrate-binding region. Met 66, which seems to induce this ‘cap’ formation, is restricted to the TIMP-1 subfamily. Thus the ‘cap’ might not be essential for complex formation in general.

Other TIMP-1 segments that have direct but weaker contacts with MMP-3 are located in the A–B loop (Val 29, Thr 33–Tyr 35, opposing MMP-3 segments S85–S86 and S154–S155) and the E–F

loop (Thr 98 and Cys 99, opposing MMP-3 segment S221–S223) of its N-terminal subdomain, and in the cross-bridged open loop (Leu 133 and Ser 134, opposing MMP-3 segment S190–S192) and the multiple-turn segment (Gln 150–Leu 152, opposing MMP-3 segment S223–S228) of its C-terminal subdomain (Figs 1 and 2). These contacts are made by means of hydrophobic side chains. The two loops from the C-terminal subdomain exhibit a relatively large mobility. The second of these TIMP-1 loops is partly undefined by electron density and flexible, like its counterpart on the MMP-3 side. This flexibility suggests that these C-terminal TIMP loops and the two opposing MMP loops merely adapt to one another, probably without making important contributions to binding energy^{6,7}. The structure further shows that the sugar chains linked to Asn 30 and Asn 78 in wild-type TIMP-1 (Fig. 2) should not collide with bound MMPs.

Superposition on the catalytic domain of full-length porcine MMP-1 (ref. 12) shows that the C-terminal domain of a full-length MMP-3

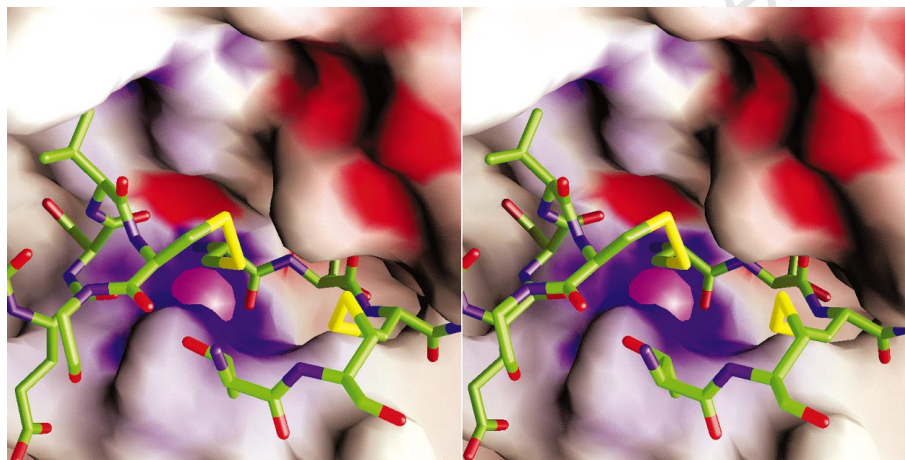


Figure 3 Stereo view from the centre of TIMP-1, towards the active-site cleft of MMP-3, which is represented by a solid surface coloured according to the electrostatic potential. The view is similar to that of Fig. 1. For simplicity, TIMP-1 (with carbons in green, oxygens in red, nitrogens in blue, and sulphurs in yellow) is represented only by the disulphide connected segments Glu 67–Gly 71 (left), Cys 1–Cys 3 (right), and Thr 98–Cys 99 (bottom segment). The N-terminal TIMP-1

residue, Cys 1, is located on top of the catalytic zinc (pink sphere), ligating it through its α -amino and its carbonyl group; Thr 2 extends towards the deep S1' pocket of MMP-3 (right). The TIMP-1 side chains of Ser 68 and Val 69 (left) nestle towards the shallow S2 subsite and the S3 pocket of MMP-3, respectively. Figure prepared using GRASP³⁰.

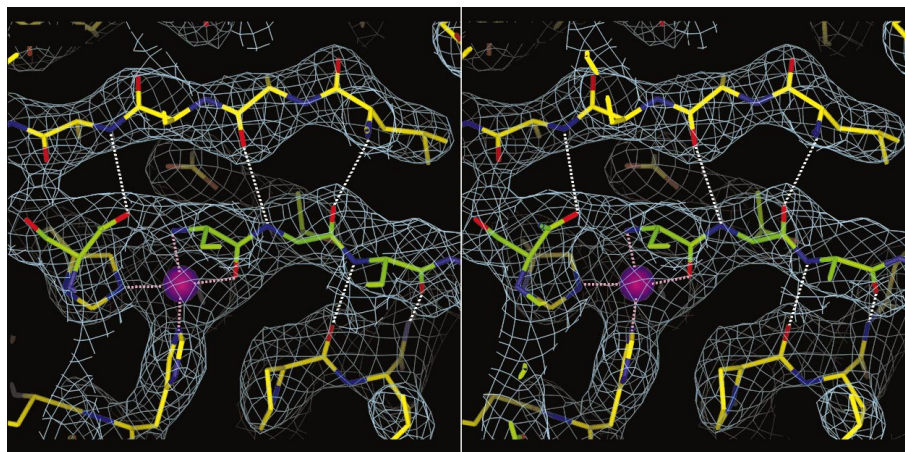


Figure 4 Stereo view of the section of the final 2.8-Å $2F_o - F_c$ electron density (blue) around the catalytic zinc (pink sphere) superimposed with the final model formed by TIMP-1 (carbon atoms green) and MMP-3 (carbon atoms yellow). The N terminus of TIMP-1 (only shown from Cys 1 to Cys 3) runs antiparallel to the MMP-3 ‘edge’ strand Leu S164–Tyr S168 (top) and parallel to segment Pro S221–Tyr S223 blocking off the specificity pocket (bottom, right)¹⁴ making six and two inter-main-chain hydrogen bonds, respectively (white dashed lines, not all of

them visible). The catalytic zinc is coordinated by the α -amino group and the carbonyl oxygen of Cys 1 (visible only up to C β) and by the three imidazole nitrogens of His S201, His S205 and His S211 (arranged anti-clockwise) in a trigonal-bipyramidal manner. The four zinc–ligand interactions visible are indicated by pink dashed lines. Contour is at 1σ and the orientation is as in Figs 1 and 3.

would just touch the bound TIMP-1, in agreement with kinetic binding studies showing that the C-terminal domain of MMP-3 contributes little to TIMP-1 binding^{6,7}. The ability of TIMP-1 to inhibit gelatinases A and B (MMP-2 and MMP-9) indicates that their additional domain (of fibronectin type II, almost certainly attached to the 'right' of the catalytic MMP domain; Fig. 1) do not impair TIMP binding. The structure of the complex between MMP-3 and TIMP-1 thus sets a limit for the protrusion of this gelatinase-specific extra domain. However, the membrane-bound MMP-14 probably exhibits an active-site region differing slightly from those of the 'classic' MMPs, as the catalytic domain of this MMP is barely inhibited by TIMP-1 (ref. 15).

The crystal structure we have described provides a foundation for the design of a new class of synthetic MMP inhibitors, based on the central two-segment docking site of TIMP-1 and the unusual zinc-chelating residue Cys 1. It also lays the grounds for rational modifications of the TIMPs to obtain more specific variants. The side chain of Thr 2 extends into the large specificity (S1') pocket of MMP-3 (refs 9–11), which is capable of accommodating much bulkier side chains. This specificity pocket differs enormously in size among MMPs, whereas residue 2 in all TIMPs known so far is Thr or Ser. Thus TIMP variants with larger side chains at position 2 might be more selective for distinct MMPs. Based on similarities in the catalytic zinc environment and chain folding, the MMPs are grouped together with other families of zinc endopeptidases, such as the serralsins, the astacins and the adamalysins (also referred to as reprotins and ADAMs¹⁶) into a larger superfamily called metzincins¹⁷. The astacins and adamalysins represent proteinases and proteins involved in morphogenetic processes, and in growth-factor activation and sperm-egg and myoblast fusion¹⁶, respectively. According to model-building studies, the serralsins and the members of the astacin family, such as the meprins or bone morphogenetic protein-1 (ref. 18), are unlikely to be inhibited by TIMP-1, owing to their much deeper and more extended active-site cleft^{19,20}. The snake-venom zinc endopeptidase adamalysin II, in contrast, should be capable of binding TIMP-1 in a similar manner to MMP-3, as they have similar surface contours around their active sites²¹. This provides the possibility of tailoring TIMPs towards more specific members of the adamalysin family, such as the recently cloned and characterized tumour-necrosis-factor- α -converting enzyme^{22,23}. Such engineered TIMPs could help provide a better understanding of the *in vivo* function of these zinc endopeptidases and may lead to therapeutic applications. □

Methods

Protein crystallization. Selenomethionine-substituted C-terminally truncated proMMP-3(Δ C) (stop codon substituted for Pro 256) was expressed in *E. coli* strain B834(DE3) (Novagen) transformed with the pET3a-proMMP-3(Δ C) vector. The recombinant SeMetproMMP-3(Δ C) was purified from inclusion bodies, refolded and activated with 1 mM 4-aminophenylmercuric acetate. The active SeMetMMP-3(Δ C) was further purified by gel filtration on Sephacryl S-200 (Pharmacia). An unglycosylated full-length TIMP-1 variant, with substitutions of Asn 30 and Asn 77 by Ala, was expressed in CHO cells transfected with the pEE14-TIMP-1 vector⁷. The unglycosylated TIMP-1 was purified on MacroPrep50Q anion exchange resin (BioRad), Green A Dye Matrex (Amicon) and Sephacryl S-200. The stoichiometric 1:1 complex of SeMetMMP-3(S83–S255) and TIMP-1 (1–184, unglycosylated) was purified by gel filtration on Sephacryl S-200. This complex was concentrated and crystallized at room temperature in about 2.5 M NaCl, 200 mM ϵ -aminocaproic acid, 20 mM Tris-HCl, 4 mM CaCl₂, 100 mM sodium acetate, pH 6.0, using the sitting-drop vapour-diffusion technique. These crystals belong to the tetragonal space group $P4_3$, have cell axes $a = b = 80.62 \text{ \AA}$, $c = 157.62 \text{ \AA}$, and contain two complexes per asymmetric unit.

Structure determination. The crystals were transferred to a cryo-buffer containing 1 M NaCl, 1 M ϵ -aminocaproic acid, 25% glycerol and 100 mM sodium acetate, pH 6.0. Diffraction data to $2.8 \text{ \AA}$ were measured at cryo-temperatures at the wiggler beamline at DORIS (DESY, Hamburg) on a 300-

mm MAR Research image plate detector. X-ray data were collected at the maximal f'' -values of the K-edges of Zn (photon energy 9,683 eV, $\sim 1.279 \text{ \AA}$) and Se (12,680 eV, about 0.979 $\text{\AA}$) determined by X-ray fluorescence spectra. At each wavelength, Bijvoet mates were scanned by 90-deg rotations from a^* to b^* and $-a^*$ to $-b^*$. These data were evaluated, merged and scaled using DENZO/SCALEPACK²⁴ yielding 24,619 independent combined reflections (corresponding to 99.6% completeness at $2.80 \text{ \AA}$, an R_{merge} of 0.11, and a multiplicity of 14.4). A pseudo-origin in the native Patterson indicated strong pseudo-I centring in the primitive tetragonal lattice. The four independent Zinc atoms were located from an anomalous difference Patterson. In parallel, position and orientation of the two independent MMP-3 catalytic domains (using an MMP-3 model kindly provided by K. Appelt, Agouron) were determined by molecular replacement using AMORE²⁵. This search yielded the correct space group and metal positions, and confirmed the presence of two independent complexes. Fourier maps calculated with phases derived from both correctly placed catalytic domains showed no significant electron density for either TIMP-1 component, however, so Bijvoet mates to $3.5 \text{ \AA}$ were merged and scaled separately. With knowledge of the metal positions and using anomalous data from the 'Zn' and the 'Zn + Se' derivatives, MAD phases were calculated using MLPHARE²⁶. Their initial mean figure of merit of 0.57 was increased to 0.79 by solvent-flattening and histogram-matching methods using DM²⁷. These protein phases were used to calculate a first $3.5\text{-\AA}$ Fourier map, which allowed tracing and model building of most parts of both TIMP-1 molecules. Both MMP-3 and TIMP-1 models were completed and refined in a cyclic manner using combined phases, the graphics programs TURBO/FRODO²⁸ and XPLOR²⁹, and by applying decreasing non-crystallographic symmetry restraints. Finally, the data were gradually extended to $2.80 \text{ \AA}$, the non-crystallographic restraints were fully removed, and only model phases were used.

Current model. The R -factor of the current model, comprising $2 \times 2,761$ active protein atoms, 473 solvent molecules, 6 calcium ions, 4 zinc ions and 4 selenium ions, is 0.197 (R_{free} 0.297) for 21,791 independent reflections for 6.00 to $2.80 \text{ \AA}$ resolution. The root-mean-squared (r.m.s.) deviation from target bond values is $0.009 \text{ \AA}$. 151 non-hydrogen atoms are currently not properly defined by electron density, in particular loop 53–56 and residues 153–154 in both TIMP-1 molecules, and the MMP-3 segment S226–S229. Omitting the seven first residues, each MMP-3 molecule shows a $0.45 \text{ \AA}$ r.m.s. deviation from the unliganded MMP-3. The defined C α -atoms of MMP-3 and TIMP-1 of complex 1 fit to their counterparts of complex 2 with r.m.s. differences of 0.35 and $0.45 \text{ \AA}$, respectively, and the relative orientation of both components to one another differs very slightly. In general, both independent complexes are similar, however, justifying a common description of their structural properties.

Received 30 April; accepted 13 June 1997

- Nagase, H. in *Zinc Metalloproteases in Health and Disease* (ed. Hooper, N. M.) 153–204 (Taylor and Francis, London, 1996).
- Coussens, L. M. & Werb, Z. Matrix metalloproteinases and the development of cancer. *Chem. Biol.* **3**, 895–904 (1996).
- Green, J. *et al.* Molecular cloning and characterization of human tissue inhibitor of metalloproteinase 4. *J. Biol. Chem.* **271**, 30375–30380 (1996).
- Willenbrock, F. & Murphy, G. Structure–function relationships in the tissue inhibitors of metalloproteinases. *Am. J. Resp. Crit. Care Med.* **150**, 5165–5170 (1994).
- Docherty, A. J. P. *et al.* Sequence of human tissue inhibitor of metalloproteinases and its identity to erythroid-potentiating activity. *Nature* **318**, 66–69 (1985).
- Murphy, G. *et al.* The N-terminal domain of human tissue inhibitor of metalloproteinases retains metalloproteinase inhibitory activity. *Biochemistry* **30**, 8097–8102 (1991).
- Huang, W. *et al.* Folding and characterization of the amino-terminal domain of human tissue inhibitor of metalloproteinase-1 (TIMP-1) expressed at high yield in *E. coli*. *FEBS Lett.* **384**, 155–161 (1996).
- Williamson, R. A. *et al.* Solution structure of the active domain of tissue inhibitor of metalloproteinases-2. A new member of the OB fold protein family. *Biochemistry* **33**, 11745–11759 (1994).
- Gooley, P. R. *et al.* NMR structure of inhibited catalytic domain of human stromelysin-1. *Nature Struct. Biol.* **1**, 111–118 (1994).
- Becker, J. W. *et al.* Stromelysin-1: Three-dimensional structure of the inhibited catalytic domain and of the C-truncated proenzyme. *Prot. Sci.* **4**, 1966–1976 (1995).
- Dhanaraj, V. *et al.* X-ray structure of a hydroxamate inhibitor complex of stromelysin catalytic domain and its comparison with members of the zinc metalloproteinase superfamily. *Structure* **4**, 375–386 (1996).
- Li, J.-Y. *et al.* Structure of full-length porcine synovial collagenase reveals a C-terminal domain containing a calcium-linked, four-bladed β -propeller. *Structure* **3**, 541–549 (1995).
- Murzin, A. G. OB-fold: common structural and functional solution for non-homologous sequences. *EMBO J.* **12**, 861–867 (1993).
- Grams, F. *et al.* X-ray structures of human neutrophil collagenase complexed with peptide hydroxamate and peptide thiol inhibitors. Implications for substrate binding and rational drug design. *Eur. J. Biochem.* **228**, 830–841 (1995).
- Will, H., Atkinson, S. J., Butler, G. S., Smyth, B. & Murphy, G. The soluble catalytic domain of

- membrane type 1 matrix metalloproteinase cleaves the propeptide of procollagenase A and initiates autocatalytic activation. *J. Biol. Chem.* **271**, 17119–17123 (1996).
16. Huovila, A. P., Ilmeida, E. A. C. & White, J. M. ADAMs and cell fusion. *Curr. Opin. Cell Biol.* **8**, 692–699 (1996).
 17. Stöcker, W. *et al.* The metzincins: Topological and sequential relations between the astacins, adamalysins, serralyisins, and matrixins (collagenases) define a superfamily of zinc-peptidases. *Protein Sci.* **4**, 823–840 (1995).
 18. Kessler, E., Takahara, K., Biniaminov, L., Brusel, M. & Greenspan, D. S. Bone morphogenetic protein-1: The type I procollagen C-proteinase. *Science* **271**, 360–362 (1996).
 19. Baumann, U., Wu, S., Flaherty, K. M. & McKay, D. B. Three-dimensional structure of the alkaline protease of *Pseudomonas aeruginosa*: a two-domain protein with a calcium binding parallel β roll motif. *EMBO J.* **12**, 3357–3364 (1993).
 20. Bode, W., Gomis-Rüth, F. X., Huber, R., Zwillig, R. & Stöcker, W. Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases. *Nature* **358**, 164–166 (1992).
 21. Gomis-Rüth, F. X., Kress, L. F. & Bode, W. First structure of a snake venom metalloproteinase: prototype for matrix metalloproteinases/collagenases. *EMBO J.* **12**, 4151–4157 (1993).
 22. Black, R. A. *et al.* A metalloproteinase disintegrin that releases tumour-necrosis factor- α from cells. *Nature* **385**, 729–733 (1997).
 23. Moss, M. L. *et al.* Cloning of a disintegrin metalloproteinase that processes precursor tumour-necrosis factor- α . *Nature* **385**, 733–736 (1997).
 24. Otwinowski, Z. & Minor, W. *DENZO: A Film Processing for Macromolecular Crystallography* (Yale University, New Haven, CT, 1993).
 25. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta crystallogr. A* **50**, 157–163 (1994).
 26. Otwinowski, Z. in *Isomorphous Replacement and Anomalous Scattering, Daresbury study weekend proceedings* (SERC Daresbury Laboratory, Warrington, 1991).
 27. Cowtan, K. *Joint CCP4 ESF-EACBM Newsletter on Protein Crystallography* Vol. 31 (1994).
 28. Roussel, A. & Cambilleau, C. *Turbo-Frodo in Silicon Graphics Geometry, Partners Directory* (Silicon Graphics, Mountain View, CA, 1989).
 29. Brünger, A. T. Crystallographic phasing and refinement of macromolecules. *Curr. Opin. Struct. Biol.* **1**, 1016–1022 (1991).
 30. Nicholls, A., Bharadwaj, R. & Honig, B. Grasp—graphical representation and analysis of surface properties. *Biophys. J.* **64**, 166 (1993).

Acknowledgements. We thank M. Braun for help with crystallization, M. T. Stubbs for reading the manuscript, and A. Lebedev and E. H. Panepucci for help with molecular replacement. This work was supported by the SFB469, the Human Capital and Mobility, and the Biotechnology programs of the European Union, the Fonds der Chemischen Industrie, the BMBF, and the NIH.

Correspondence and requests for materials should be addressed to W.B. (e-mail: bode@biochem.mpg.de). The coordinates have been deposited at the Protein Data Bank (accession no. PDB ID code IUEA).

ER-to-Golgi transport visualized in living cells

John F. Presley, Nelson B. Cole, Trina A. Schroer*, Koret Hirschberg, Kristien J. M. Zaal & Jennifer Lippincott-Schwartz

Cell Biology and Metabolism Branch, National Institute of Child Health and Human Development, Building 18T, NICHD, NIH, Bethesda, Maryland 20892, USA

* Department of Biology, The Johns Hopkins University, Baltimore, Maryland 21218, USA

Newly synthesized proteins that leave the endoplasmic reticulum (ER) are funnelled through the Golgi complex before being sorted for transport to their different final destinations. Traditional approaches have elucidated the biochemical requirements for such transport^{1–3} and have established a role for transport intermediates^{4–8}. New techniques for tagging proteins fluorescently^{9,10} have made it possible to follow the complete life history of single transport intermediates in living cells, including their formation, path and velocity *en route* to the Golgi complex. We have now visualized ER-to-Golgi transport using the viral glycoprotein ts045 VSVG tagged with green fluorescent protein (VSVG-GFP). Upon export from the ER, VSVG-GFP became concentrated in many differently shaped, rapidly forming pre-Golgi structures, which translocated inwards towards the Golgi complex along microtubules by using the microtubule minus-end-directed motor complex of dynein/dynactin. No loss of fluorescent material from pre-Golgi structures occurred during their translocation to the Golgi complex and they frequently stretched into tubular shapes. Together, our results indicate that these pre-Golgi carrier structures moving unidirectionally along microtubule tracks are

responsible for transporting VSVG-GFP through the cytoplasm to the Golgi complex. This contrasts with the traditional focus on small vesicles as the primary vehicles for ER-to-Golgi transport.

VSVG protein from the ts045 mutant strain of vesicular stomatitis virus has been widely used to study membrane transport because of its reversible misfolding and retention in the ER at 40°C and its ability to move out of the ER and into the Golgi complex upon temperature reduction to 32°C^{11–13}. Green fluorescent protein (GFP)^{9,10} was attached to the cytoplasmic tail of VSVG ts045. When expressed in COS cells incubated at 40°C, VSVG-GFP fluorescence was localized exclusively to the ER (Fig. 1a; 40°C), a result consistent with the temperature-sensitive folding phenotype of the viral glycoprotein. Upon temperature shift to 32°C for 15 min the fusion protein redistributed into the juxtanuclear Golgi complex (Fig. 1a; 32°C) and in biochemical assays¹⁴ became resistant to endo H (data not shown). This time frame for Golgi transport and processing of the fusion protein was indistinguishable from that of untagged VSVG protein in COS cells and was followed by redistribution of VSVG-GFP to the cell surface. Thus, neither the GFP tag nor the imaging of living cells appeared to interfere with the normal folding and transport properties of VSVG.

The initial targets for vesicles budding from the ER *en route* to the Golgi complex are pleiomorphic tubulovesicular structures found in the Golgi region and in peripheral sites^{4–8}, which are enriched in several proteins (including the COPI subunit β -COP and ERGIC53)^{15–18}, and whose proposed function is to concentrate and sort secretory cargo^{19,20}. At reduced temperatures (such as 15°C) these pre-Golgi structures become markedly enlarged and accumulate secretory products²¹, presumably owing to a rate-limiting step in membrane transport through these intermediates^{22,23}. VSVG-GFP accumulated in such intermediate structures upon incubation for 3 h at 15°C, co-localizing extensively with β -COP (Fig. 1a; 15°C) and ERGIC53 (data not shown).

To examine how VSVG-GFP is transported from such intermediates into the Golgi region, cells expressing the chimaeric proteins at 15°C were placed on a microscope stage warmed to 32°C and fluorescent images were collected at 3.6-s intervals. As shown in Fig. 1b and c (also see Quicktime movie at <http://dir.nichd.nih.gov/CBMB/pb1labob.html>), peripheral pre-Golgi structures containing VSVG-GFP translocated rapidly as units into the centrosomal region where they merged into the large fluorescent area marking the Golgi complex. Figure 1b (centre) maps the movement of several of these structures over the course of 9 min upon warm-up from 15°C to 32°C. Such structures moved along straight or curvilinear paths towards the cell centre at speeds of up to 1.4 $\mu\text{m s}^{-1}$. Individual pre-Golgi elements were often greater than 1.5 μm in diameter and frequently changed shape during translocation. Staining of fixed cells with β -COP antibodies during this period revealed the continued association of β -COP and the VSVG-GFP-containing pre-Golgi structures as they clustered inwards to the Golgi complex (data not shown). Inward translocation of two pre-Golgi structures (marked by arrows) is shown in Fig. 1c. A single long tubule was often seen extending in the direction of motion of the pre-Golgi element (Fig. 1b, inset), suggesting that a motor on the tip of the tubule was pulling the mass of pre-Golgi membranes towards the Golgi complex.

Pre-Golgi structures became motile at variable times after warm-up to 32°C and moved in a stop-and-go fashion toward the Golgi complex. The velocity and path of a single element containing VSVG-GFP, which illustrates this motion, is plotted in Fig. 2a and b. Pre-Golgi structures did not lose fluorescence intensity as they translocated inwards (Fig. 2c), indicating that VSVG-GFP molecules present within them did not bud off into small vesicle carriers, but were carried *en masse* with this structure into the Golgi region.

Cells expressing VSVG-GFP were also visualized upon shifting the temperature directly from 40°C to 32°C (Fig. 3a–d and Quicktime movie). Within 1 to 5 min of the shift to lower

temperature, numerous bright fluorescent structures began to appear at widely dispersed sites within the reticular ER labelled by VSVG-GFP (Fig. 3a, arrows). These structures were smaller than the pre-Golgi intermediates observed at 15 °C, but like the 15 °C structures were labelled by antibodies against β -COP (data not shown). When such structures formed, their relative fluorescent intensity was as much as 4 to 10 times above the diffuse fluorescence in the ER (see inset, Fig. 3a). Imaging of the lifetime of these structures revealed that they formed within seconds and stayed at peripheral sites for only brief periods before translocating into the Golgi region (Fig. 3c). Multiple motile structures rarely, if ever, originated sequentially from the same single site but seemed to arise randomly within the ER. Formation and translocation of VSVG-

GFP-enriched structures repeatedly occurred until virtually all of the ER-derived fluorescence was redistributed into the Golgi region 15 min after temperature shift (Fig. 3b).

Peripheral pre-Golgi structures containing VSVG-GFP that appeared upon temperature shift from 40 °C to 32 °C translocated rapidly into the Golgi region in a 'stop-and-go' manner, following inward trajectories similar to those of the larger pre-Golgi structures observed during warm-up from 15 °C (Fig. 3c). Movement was not synchronous. At any one time there was a small population of immobile structures. Once a peripheral structure began to move, its fluorescence intensity remained constant. Maximum velocities for movement were around $1.4 \mu\text{m s}^{-1}$, similar to velocities of pre-Golgi elements in cells warmed from 15 °C to 32 °C, and consistent

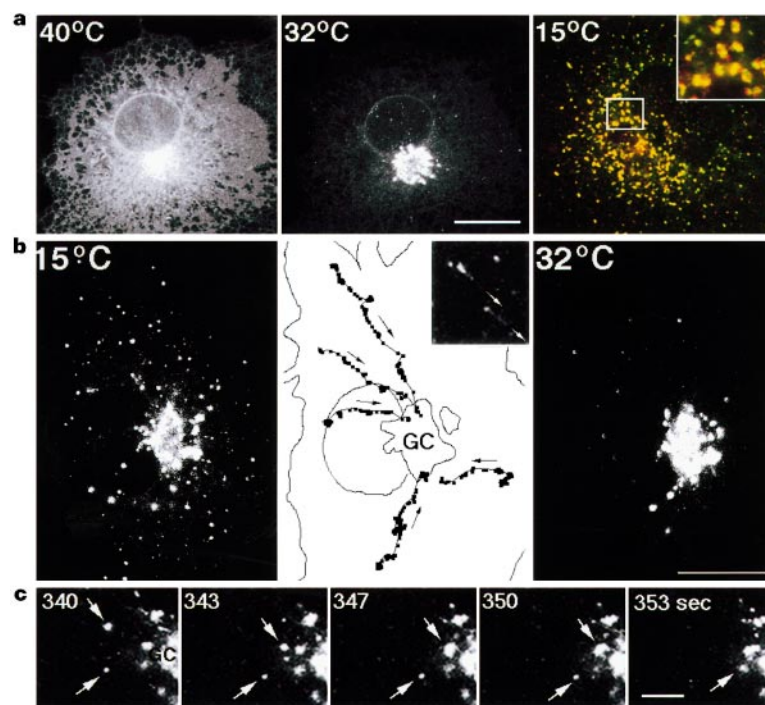


Figure 1 Expression and transport of VSVG-GFP in living cells. **a**, VSVG-GFP expressing COS cells were incubated at 40 °C for 12 h (left panel, 40 °C) and then shifted to 32 °C for 15 min (middle panel, 32 °C) or to 15 °C for 3 h (right panel, 15 °C; fixed (2% formaldehyde) and stained with antibodies to β -COP). The left and middle panels show fluorescence images taken of the same living cell. Green staining in right panel is VSVG-GFP, red staining is β -COP and yellow represents overlap of the two. Scale bar, 20 μm . **b**, VSVG-GFP-expressing COS cells incubated at 15 °C for 3 h and fluorescence images collected at 32 °C. Left panel shows the distribution of VSVG-GFP at the time of warm-up, and right panel shows the distribution 9 min later in the same cell. Paths of pre-Golgi structures are shown in the middle panel (3.5 s intervals). Arrows point in the direction of motion. Inset of middle panel shows high magnification of motile pre-Golgi structure extending a tubule in its direction of movement (marked by arrows). Scale bar, 20 μm . **c**, Image series showing movement of two pre-Golgi structures (see upper and lower arrows) into the Golgi region (GC). Scale bar, 6 μm . (A Quicktime movie sequence from the experiment in **b** is available at <http://dir.nichd.nih.gov/CBMB/pb11a-bob.html>)

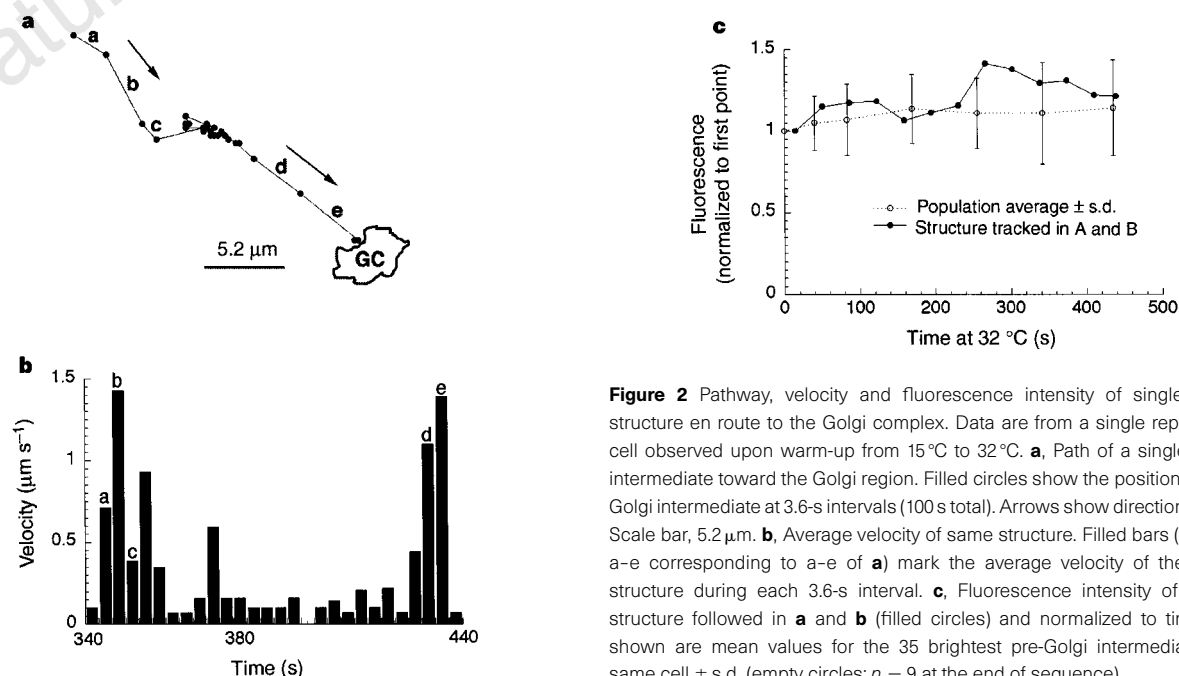


Figure 2 Pathway, velocity and fluorescence intensity of single pre-Golgi structure en route to the Golgi complex. Data are from a single representative cell observed upon warm-up from 15 °C to 32 °C. **a**, Path of a single pre-Golgi intermediate toward the Golgi region. Filled circles show the position of the pre-Golgi intermediate at 3.6-s intervals (100 s total). Arrows show direction of motion. Scale bar, 5.2 μm . **b**, Average velocity of same structure. Filled bars (with letters a-e corresponding to a-e of **a**) mark the average velocity of the pre-Golgi structure during each 3.6-s interval. **c**, Fluorescence intensity of the same structure followed in **a** and **b** (filled circles) and normalized to time 0. Also shown are mean values for the 35 brightest pre-Golgi intermediates in the same cell $\pm$ s.d. (empty circles; $n = 9$ at the end of sequence).

with maximal velocities of microtubule-based motors²⁴. Extension of tubule processes from a large proportion of the pre-Golgi structures (observed either through the eyepiece of the microscope or in confocal images) indicated these structures were pleiomorphic intermediates rather than single small vesicles (which are incapable of such shape changes) (Fig. 3d). Because virtually every VSVG-GFP-containing pre-Golgi intermediate could be tracked into the Golgi region as an intact unit within a few minutes of its formation (with no significant change in relative fluorescent intensity), VSVG-GFP was carried *en bloc* by these structures to the Golgi complex, as during warm-up from 15 °C. Similar results were found in HeLa, NRK, MDCK, CHO and primary glial cells transfected with VSVG-GFP.

The fate of pre-Golgi structures upon translocation to the Golgi region was observed in cells whose Golgi area was emptied of fluorescence by photobleaching²⁵ (Fig. 3e and Quicktime movie). Within 5 s to 1 min after photobleaching, pre-Golgi structures began translocating and delivering their contents to the Golgi complex. Within 3 to 5 min of photobleaching, the profile of VSVG-GFP-containing Golgi membranes was similar to the pre-bleach image and there appeared to be no other source than these pre-Golgi structures for delivery of fluorescence. Possible mechanisms for transfer of fluorescence from pre-Golgi structures to the Golgi complex include vesicle budding and fusion, or fusion of whole pre-Golgi structures to the *cis* face of the Golgi complex. No pre-Golgi structures were observed moving back to the cell periphery

with any portion of their cargo of VSVG-GFP.

In cells treated with nocodazole to depolymerize microtubules at the time of temperature shift to 32 °C, VSVG-GFP was exported out of the ER into peripheral pre-Golgi elements (Fig. 4a–c). These structures, however, failed to acquire directed motility and remained stationary in the cell periphery (Fig. 4e). Antibody staining of these structures in fixed cells revealed that they contained β -COP and lacked Golgi enzymes (data not shown). Moreover, no significant processing of VSVG-GFP to forms resistant to endo H was observed under these conditions (with less than 5% VSVG-GFP resistant to endo H after 20 min). These results indicate that microtubules are required for translocation of pre-Golgi structures inwards to the cell centre where the Golgi complex normally resides. Upon removal of nocodazole from cells, the peripheral pre-Golgi structures began to move and translocated as individual aggregates following curvilinear trajectories into the cell centre (Fig. 4f).

Over time in cells treated with nocodazole at 32 °C, peripheral pre-Golgi elements containing VSVG-GFP grew increasingly brighter and larger in size (Fig. 4d). This was in contrast to pre-Golgi structures in cells with intact microtubules, whose fluorescence intensity remained constant as they translocated into the Golgi region (Fig. 2c). This difference suggests that pre-Golgi structures in nocodazole-treated cells remain closely associated with ER exit sites where they can continually receive VSVG-GFP and other proteins exported from the ER. Over 1 to 2 h in nocodazole-treated cells, Golgi enzymes slowly redistributed into

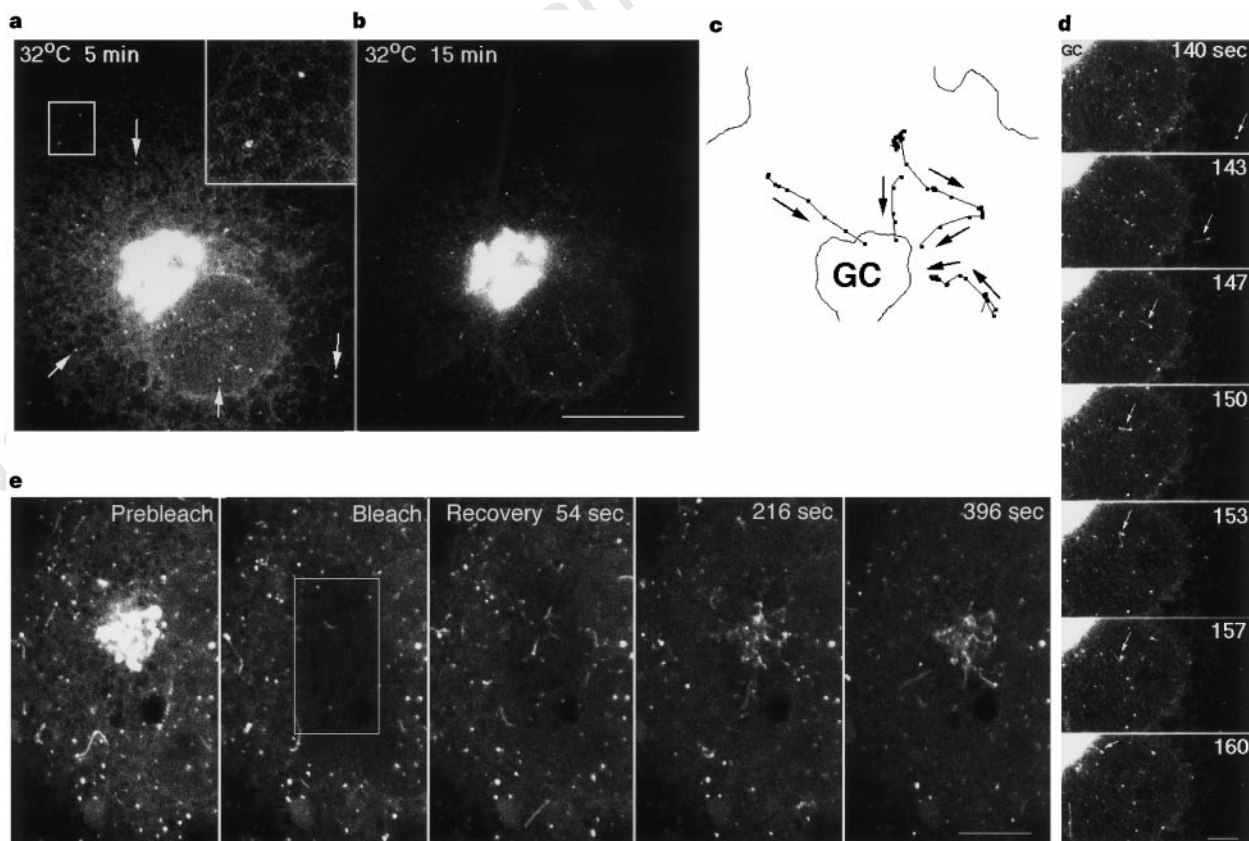


Figure 3 ER to Golgi transport of VSVG-GFP visualized upon shift from 40 °C to 32 °C or in cells whose Golgi area is photobleached. **a–d**, VSVG-GFP-expressing COS cells were incubated for 12 h at 40 °C and then shifted to 32 °C. **a**, Distribution of VSVG-GFP after 5 min at 32 °C. Arrows show examples of pre-Golgi intermediates. Inset shows two pre-Golgi structures with fluorescence intensity several fold greater than the ER. **b**, Same cell after 15 min at 32 °C. Scale bar, 20 μ m. **c**, Paths taken by several representative pre-Golgi structures en route to the Golgi complex (same cell as in **a** and **b**). Positions shown are at 8.6-s intervals.

d, Image series showing shape change of a pre-Golgi intermediate as it translocated to the Golgi complex over the time interval shown. Scale bar, 4 μ m. **e**, VSVG-GFP-expressing COS cells were incubated for 12 h at 40 °C, shifted to 15 °C for 3 h and then warmed to 32 °C. Fluorescence associated with the Golgi complex was photobleached with high-intensity laser light and subsequent inward delivery of VSVG-GFP from pre-Golgi intermediates followed over the time periods indicated. Scale bar, 9.6 μ m.

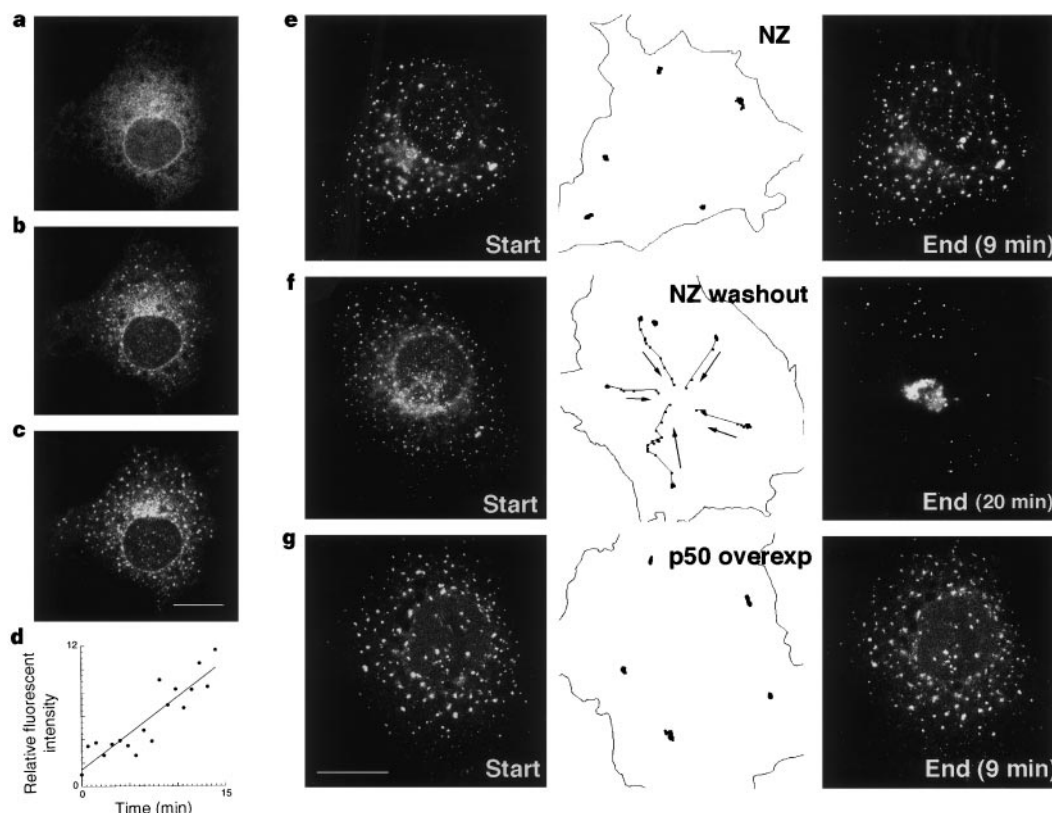


Figure 4 Role of microtubules and dynein/dynactin in translocation of pre-Golgi structures. **a–c**, COS cells expressing VSVG-GFP at 40°C were placed on ice for 15 min with 1 µg ml⁻¹ nocodazole, and warmed to 32°C in the presence of drug. Images were collected at 0 min (**a**), 10 min (**b**) and 20 min (**c**) at 32°C. Scale bar, 20 µm. **d**, Fluorescence intensity of pre-Golgi structure in nocodazole-treated cell from image sequences collected at 3.6-s intervals at 32°C. **e–g**, Path of pre-Golgi structures in cells in which VSVG-GFP was accumulated in the ER at 40°C and then images taken at 10-s intervals at 32°C under different conditions. 'Start' shows the distribution of VSVG-GFP at the beginning of the sequence, 'End' shows the distribution 9 or 20 min later. Middle panels show the path of structures tracked over the period between 'start' and 'end'. **e** shows a 9-min sequence of a cell treated with nocodazole as in **a** and images taken starting after 20 min in the drug. **f** shows a 20-min sequence of a cell treated with nocodazole as in **e** and then washed free of the drug. **g**, A 9-min sequence of a cell co-expressing VSVG-GFP and dynamin beginning after 10 min at 32°C. A Quicktime movie sequence of **e–g** is available (see Fig. 1 legend).

the peripheral structures containing VSVG-GFP (data not shown), leading to reformation of Golgi stacks at ER exit sites as previously described¹⁴.

The loss of motility of pre-Golgi elements after microtubule disruption with nocodazole suggested that a microtubule motor protein powers their inward movement. Consistent with this, energy depletion with deoxyglucose and azide blocked their inward translocation (data not shown). Cytoplasmic dyneins are minus-end-directed, microtubule-dependent ATPases involved in organelle motility. At least one isoform of cytoplasmic dynein³¹ (dynein 1) interacts with dynactin^{26,27}, another multisubunit complex, to target subcellular organelles to microtubules. To inhibit dynein activity, we overexpressed the p50/dynactin subunit of dynactin in VSVG-GFP-expressing cells. Overexpression of dynactin disrupts the dynactin complex and dissociates both it and cytoplasmic dynein from mitotic kinetochores²⁸. It also results in dispersion of the Golgi complex, presumably because of a role of dynein/dynactin in Golgi positioning²⁹. As shown in Fig. 4g, VSVG-GFP-containing peripheral structures failed to translocate into the centrosomal region upon shift from 40°C to 32°C in cells overexpressing dynactin and remained as stationary structures in the cell periphery even though microtubules were intact. This inhibitory effect occurred before the effects of dynactin on Golgi fragmentation (not shown) and suggested that translocation of peripheral pre-Golgi structures inwards along microtubules is powered by the microtubule motor complex of dynein/dynactin.

In summary, use of VSVG-GFP chimaeras to follow membrane

transport in living cells has revealed new information about the nature and lifetime of transport intermediates involved in membrane traffic between the ER and Golgi complex. VSVG-GFP rapidly emerged from the ER at multiple peripheral sites, resulting in the accumulation of chimaeric protein in numerous peripheral membrane structures. These structures translocated as units along microtubules into the Golgi region without loss of accumulated VSVG-GFP, and were often large enough to deform in shape during transport, suggesting they are non-vesicular membrane transport intermediates. The data further underscore the importance of microtubules and microtubule-based motors: pre-Golgi transport intermediates required microtubules to translocate inwards to the Golgi complex and appeared to use the dynein/dynactin motor complex for this purpose. □

Methods

Cloning and expression of VSVG-GFP. VSVG-GFP was generated by fusing enhanced GFP (Clontech, Palo Alto, CA) directly after the final amino acid of the carboxy terminus of VSVG ts045 (ref. 30; provided by J. Rose) by standard two-stage polymerase chain reaction (PCR) cloning methods. The construct was cloned into the CMV promoter-driven expression plasmid, pCDM8.1 (Invitrogen), and transiently expressed in COS cells by electroporation.

Fluorescence imaging. GFP-expressing cells were imaged at 32°C in buffered medium with a Zeiss LSM 410 confocal microscope system having a ×100 Zeiss planapo objective (NA1.4) or with a microscope equipped with a cooled charge-coupled device (CCD) system described in ref. 14. Detectors on both microscopes gave a linear response to light under the conditions in this study.

Small 50–100 nm structures (vesicles) can easily be detected with our imaging system given the presence of enough fluorescent signal (we can visualize coated pits or caveolae using fluorescent antibody labelling). Line averaging rather than frame averaging was used in confocal imaging (images were acquired with a single top-to-bottom scan) to prevent artefacts caused by movement during data acquisition. Temperature was controlled using a Nevtex air stream stage incubator (Bursville). The GFP molecule was excited with the 488 line of a krypton-argon laser and the images made with a 515–540 bandpass filter. Images were processed using NIH image software (Wayne Rasband Analytics, NIH) and were printed with a Fujix Pictography 3000 digital printer.

Fluorescence intensity of pre-Golgi intermediates formed at 15°C was measured by centering on a circular region of interest 6 pixels in diameter (1 pixel = 0.26 µm in the source images for Fig. 2) on the brightest point of the punctate pre-Golgi intermediate in each image and measuring the total fluorescence in this region. If this region was insufficient to surround the apparent area of the pre-Golgi intermediate in the first frame completely, a larger region was used for the first and kept constant for subsequent frames. A nearby background region of equal area was similarly measured and subtracted. Background pixel values were <20% of the average intensity of these structures. Fluorescence intensity was normalized to the first time point.

Received 21 March; accepted 16 June 1997.

- Rothman, J. E. & Wieland, F. T. Protein sorting by transport vesicles. *Science* **272**, 227–234 (1996).
- Schekman, R. & Orci, L. Coat proteins and vesicle budding. *Science* **271**, 1526–1533 (1996).
- Aridor, M., Bannykh, S., Rowe, T. & Balch, W. E. Sequential coupling between CopII and CopI vesicle coats in endoplasmic reticulum to Golgi transport. *J. Cell Biol.* **131**, 1–19 (1995).
- Plutner, H., Davidson, H. W., Saraste, J. & Balch, W. E. Morphological analysis of protein transport from the ER to Golgi membranes in digitonin-permeabilized cells: role of the p58 containing compartment. *J. Cell Biol.* **119**, 1097–1116 (1992).
- Saraste, J. & Svensson, K. Distribution of the intermediate elements operating in ER to Golgi transport. *J. Cell Sci.* **100**, 415–430 (1991).
- Saraste, J. & Kuismanen, E. Pathways of protein sorting and membrane traffic between the rough endoplasmic reticulum and the Golgi complex. *Semin. Cell Biol.* **3**, 343–355 (1992).
- Krijnse-Locker, J., Ericsson, M., Rottier, P. J. & Griffiths, G. Characterization of the budding compartment of mouse hepatitis virus: Evidence that transport from the RER to the Golgi complex requires only one vesicular transport step. *J. Cell Biol.* **124**, 55–70 (1994).
- Stinchcombe, J. C., Nomoto, H., Cutler, D. F. & Hopkins, C. R. Anterograde and retrograde traffic between the rough endoplasmic reticulum and the Golgi complex. *J. Cell Biol.* **131**, 1387–1401 (1995).
- Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G. & Cormier, M. J. Primary structure of the Aequorea victoria green-fluorescent protein. *Gene* **111**, 229–233 (1992).
- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science* **263**, 802–805 (1994).
- Kreis, T. E. & Lodish, H. F. Oligomerization is essential for transport of vesicular stomatitis viral glycoprotein to the cell surface. *Cell* **46**, 929–937 (1986).
- Beckers, C. J., Keller, D. S. & Balch, W. E. Semi-intact cells permeable to macromolecules: Use in reconstitution of protein transport from the endoplasmic reticulum to the Golgi complex. *Cell* **50**, 523–534 (1987).
- Bergmann, J. E. Using temperature-sensitive mutants of VSV to study membrane protein biogenesis. *Methods Cell Biol.* **32**, 85–110 (1989).
- Cole, N. B., Scialy, N., Marotta, A., Song, J. & Lippincott-Schwartz, J. Golgi dispersal during microtubule disruption: regeneration of Golgi stacks at peripheral endoplasmic reticulum exit sites. *Mol. Biol. Cell* **7**, 631–650 (1996).
- Schweizer, A., Fransen, J. A. M., Bachi, T., Ginsel, L. & Hauri, H.-P. Identification, by a monoclonal antibody, of a 53 kD protein associated with a tubulo-vesicular compartment at the cis-side of the Golgi apparatus. *J. Cell Biol.* **107**, 1643–1653 (1988).
- Lippincott-Schwartz, J., Cole, N. B., Marotta, A., Conrad, P. A. & Bloom, G. S. Kinesin is the motor for microtubule-mediated Golgi-to-ER membrane traffic. *J. Cell Biol.* **128**, 293–306 (1995).
- Pepperkok, R. et al. βCOP is essential for biosynthetic membrane transport from the endoplasmic reticulum to the Golgi complex *in vivo*. *Cell* **74**, 71–82 (1993).
- Peter, F., Plutner, H., Zhu, H., Kreis, T. E. & Balch, W. E. β-COP is essential for transport of protein from the endoplasmic reticulum to the Golgi *in vitro*. *J. Cell Biol.* **122**, 1155–1168 (1993).
- Balch, W. E., McCaffery, J. M., Plutner, H. & Farquhar, M. G. Vesicular stomatitis virus is sorted and concentrated upon exit from the endoplasmic reticulum. *Cell* **76**, 841–852 (1994).
- Bannykh, S. I., Rowe, T. & Balch, W. E. The organization of endoplasmic reticulum export complexes. *J. Cell Biol.* **135**, 19–35 (1996).
- Kuismanen, E. & Saraste, J. Low temperature-induced transport blocks as tools to manipulate membrane traffic. *Methods Cell Biol.* **32**, 257–274 (1989).
- Hauri, H.-P. & Schweizer, A. The endoplasmic reticulum–Golgi intermediate compartment. *Curr. Opin. Cell Biol.* **4**, 600–608 (1992).
- Lotti, L. V., Torrisi, M. R., Pascale, M. C. & Bonatti, S. Immunocytochemical analysis of the transfer of vesicular stomatitis virus G glycoprotein from the intermediate compartment to the Golgi complex. *J. Cell Biol.* **118**, 43–50 (1992).
- Walker, R. A. & Sheetz, M. P. Cytoplasmic microtubule-associated motors. *Annu. Rev. Biochem.* **62**, 429–451 (1993).
- Cole, N. B. et al. Diffusional mobility of Golgi proteins in membranes of living cells. *Science* **273**, 797–801 (1996).
- Schroer, T. A., Bingham, J. B. & Gill, S. R. Actin-related protein 1 and cytoplasmic dynein-based motility: What's the connection? *Trends Cell Biol.* **6**, 212–215 (1996).
- Gaglio, T. et al. Opposing motor activities are required for the organization of the mammalian mitotic spindle pole. *J. Cell Biol.* **135**, 399–414 (1996).
- Echeverri, C. J., Paschal, B. B., Vaughan, K. T. & Vallee, R. B. Molecular characterization of the 50-kD subunit of dynactin reveals function for the complex in chromosome alignment and spindle organization during mitosis. *J. Cell Biol.* **132**, 617–633 (1996).
- Burkhardt, J. K., Echeverri, C. J. & Vallee, R. B. Overexpression of the p50 subunit of dynactin perturbs the positioning of the Golgi apparatus and endosomes. *Mol. Biol. Cell* **6**, 266a (1995).

- Gallione, C. J. & Rose, J. K. A single amino acid substitution in a hydrophobic domain causes temperature-sensitive cell-surface transport of a mutant viral glycoprotein. *J. Virol.* **54**, 374–382 (1985).
- Vaisberg, E. A., Grissom, P. M. & McIntosh, J. R. Mammalian cells express three distinct dynein heavy chains that are localized to different cytoplasmic organelles. *J. Cell Biol.* **133**, 831–842 (1996).

Acknowledgements. We thank R. Klausner, E. Siggia, J. Bonifacio, C. Smith, J. Donaldson, J. Ellenberg and R. Stearnman for valuable comments and suggestions, and J. Rose for his generous gift of reagent.

Correspondence and requests for materials should be addressed to J.L.-S. (e-mail: jlipin@helix.nih.gov).

Smad4 and FAST-1 in the assembly of activin-responsive factor

Xin Chen, Ellen Weisberg, Valerie Fridmacher, Minoru Watanabe, Grace Naco & Malcolm Whitman

Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115-5730, USA

Members of the TGF-β superfamily of signalling molecules work by activating transmembrane receptors with phosphorylating activity (serine–threonine kinase receptors)¹; these in turn phosphorylate and activate² SMADs^{3,4}, a class of signal transducers. Activins are growth factors that act primarily through Smad2^{5–7}, possibly in partnership with Smad4, which forms heteromeric complexes with different ligand-specific SMADs after activation^{8,9}. In frog embryos, Smad2 participates in an activin-responsive factor (ARF), which then binds to a promoter element of the *Mix.2* gene¹⁰. The principal DNA-binding component of ARF is FAST-1 (ref. 11), a transcription factor with a novel winged-helix structure. We now report that Smad4 is present in ARF, and that FAST-1, Smad4 and Smad2 co-immunoprecipitate in a ligand-regulated fashion. We have mapped the site of interaction between FAST-1 and Smad2/Smad4 to a novel carboxy-terminal domain of FAST-1, and find that overexpression of this domain specifically inhibits activin signalling. In a yeast two-hybrid assay, the FAST-1 carboxy terminus interacts with Smad2 but not Smad4. Deletion mutants of the FAST-1 carboxy terminus that still participate in ligand-regulated Smad2 binding no longer associated with Smad4 or ARF. These results indicate that Smad4 stabilizes a ligand-stimulated Smad2–FAST-1 complex as an active DNA-binding factor.

Smad2 associates in a ligand-regulated manner with another member of the Smad family, Smad4 (ref. 8). To test whether Smad4 is a component of ARF, we expressed epitope-tagged Smad4 by injection of messenger RNA into 2-cell embryos either with or without co-injection of activin mRNA; we then tested whether ARF isolated from these embryos at stage 9 could be supershifted in an electrophoretic mobility-shift assay (EMSA) by incubation with anti-tag antibody (Fig. 1a). In this assay, Smad4 was incorporated into ARF, but overexpression of Smad4 was insufficient to stimulate ARF formation in the absence of ligand activation (Fig. 1a, lanes 2 and 3).

Incorporation of Smad4 into ARF could reflect either the association of Smad2, Smad4 and FAST-1 into the same complex, or the formation of two distinct types of ARF: Smad2-containing ARF and Smad4-containing ARF. To distinguish between these possibilities, we overexpressed Smad2 bearing a Myc-epitope tag and Smad4 bearing a haemagglutinin (HA) epitope tag, and did an EMSA supershift with each anti-tag antibody alone or together. Addition of both anti-tag antibodies supershifted ARF-containing tagged Smad4 and Smad2 to a position higher than either antibody alone (Fig. 1b), indicating that Smad2 and Smad4 are present in the same complex, rather than in two discrete subsets of ARF complexes.

Identical results were obtained when tagged Smad3 was expressed rather than Smad2, indicating that Smad3 can also participate in ARF in an activin-regulated manner (results not shown).

To test whether the Smad MH2 domain is responsible for the Smad2 association with ARF, we tested overexpressed, tagged Smad2 MH2 domain for incorporation into the ARF complex by using supershift EMSA (Fig. 1c). Smad2 MH2 domain formed an activin-response element (ARE)-binding complex that migrated at a lower position than endogenous ARF, presumably because the MH2 domain is smaller than full-length Smad2; formation of this complex was stimulated by activin (Fig. 1c, lanes 2 and 4). This faster migrating complex contained Smad4, because it was supershifted by antibody directed against a tag on Smad4 (Fig. 1c, lane 9).

ARF is detected in EMSA as a complex with its DNA target ARE. To examine the association of domains of FAST-1 with Smad2 and Smad4 separately from the formation of the ARF-ARE complex, we tested whether FAST-1, Smad2 and Smad4 could be co-immunoprecipitated from embryonic lysates. Lysates from embryos expres-

sing GST-tagged FAST-1 and Myc-tagged Smad2 either with or without activin stimulation were precipitated with anti-GST antibody, and the immunoprecipitates blotted with anti-Myc antibody. Under these conditions, tagged Smad2 precipitates with FAST-1, an association that is increased by ligand stimulation (Fig. 2a). Co-expression of tagged Smad1 with FAST-1 does not, however, lead to detectable co-precipitation of Smad1 (Fig. 2a). In the absence of tagged FAST expression, Smad2 does not precipitate; equal expression of Smad2 and Smad1 was confirmed by direct blotting of a portion of the embryonic lysate (Fig. 2a, lower panel). Similar results were obtained upon immunoprecipitating Smad2 and immunoblotting to detect the tag on FAST-1 (not shown). Co-precipitation of FAST-1 and Smad2 was clearly detectable in the absence of activin stimulation. This ligand-independent association was not reflected in EMSA as an ARE-binding activity that was supershiftable with anti-Myc antibody (X.C., unpublished results). Co-immunoprecipitation therefore appears to identify not only an activin-dependent complex of FAST-1 and Smad2, but also an

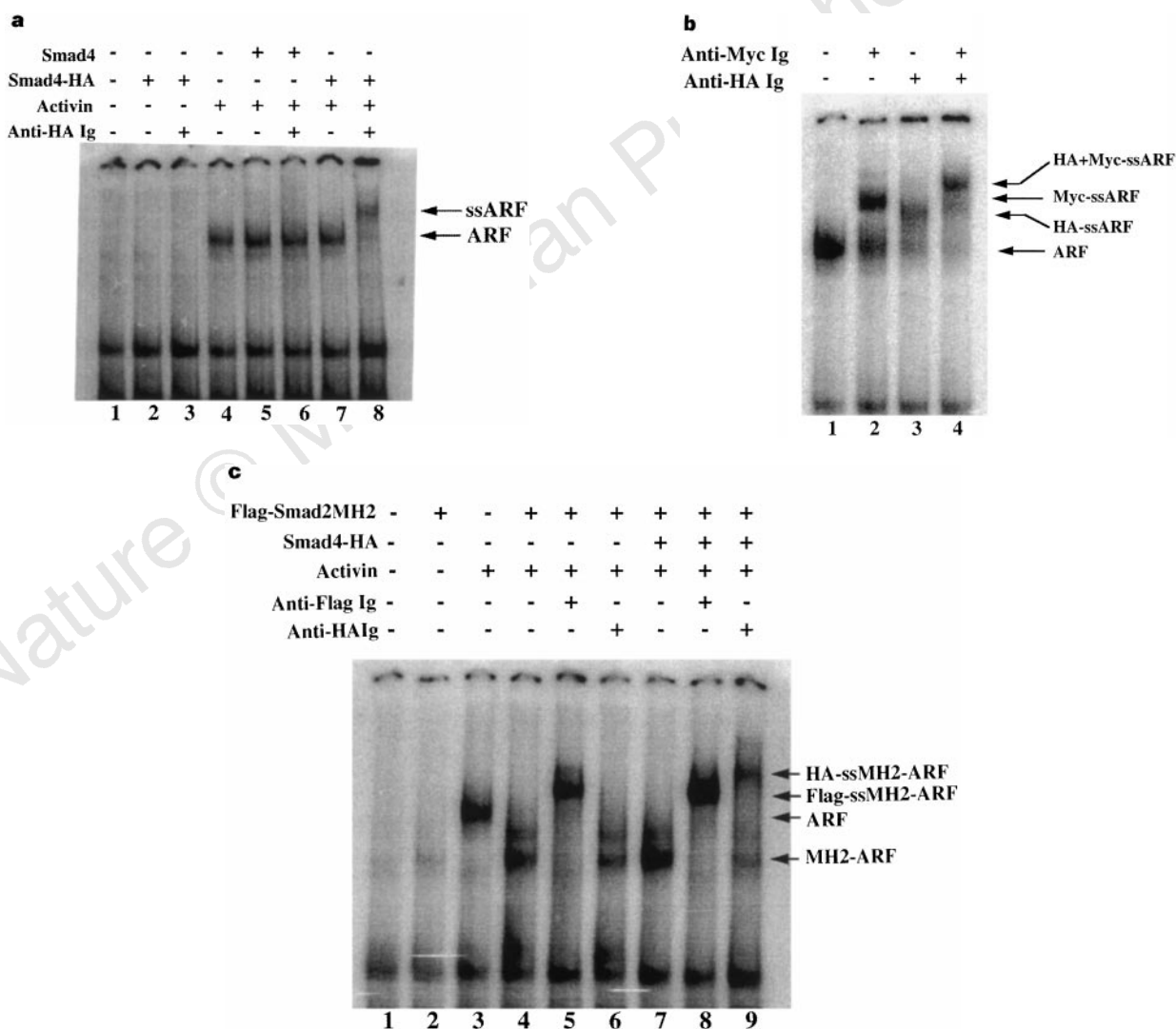


Figure 1 Molecular composition of ARF. Incorporation of tagged SMAD family constructs into ARF was tested by supershift EMSA. **a**, Incorporation of Smad4 into ARF. Tagged (lanes 2, 3, 7, 8) or untagged (lanes 5, 6) Smad4 was expressed in embryos; incorporation of HA-tagged protein into ARF was tested by incubation with anti-HA antibody (lanes 3, 6, 8). ssARF (supershift ARF) is antibody-bound ARF. **b**, Presence of Smad2 and Smad4 in the same complex. Myc-tagged Smad2 and HA-tagged Smad4 were expressed in embryos with activin and assayed by supershift EMSA. HA-ssARF, ARF complexes with only anti-HA antibody bound;

Myc-ssARF, complexes with only anti-Myc antibody bound; HA + Myc-ssARF complexes with shifted mobility due to the binding of both antibodies. **c**, Incorporation of Smad2-MH2 with Smad4 into ARF. Flag-tagged Smad2-MH2 and HA-tagged Smad4 were expressed and assayed for ARF incorporation. MH2-ARF, ARF migrating with increased mobility owing to the incorporation of Smad2-MH2; Flag-ssMH2-ARF, Flag-Smad2-MH2-ARF complex supershifted with anti-Flag; HA-ssMH2-ARF, Smad2-MH2-ARF complex supershifted by anti-HA (recognizing Smad4 in the complex).

activin-independent complex that is either not competent for DNA binding or unstable under the conditions of our EMSA.

To investigate Smad4 precipitation with FAST-1, Myc-tagged Smad4 was co-expressed with GST-tagged FAST-1 with or without stimulation by activin. Smad4 associated with FAST-1 in the presence of activin stimulation, but not in its absence (Fig. 2b). As previously reported⁸, Smad4 co-precipitation with Smad2 was also activin-stimulated (not shown). Results were identical when tagged Smad4 was immunoprecipitated and blotted for tagged FAST-1, or when Flag-tagged FAST-1 rather than GST-FAST-1 was used (not shown).

The DNA-binding component of ARF, FAST-1, contains a predicted winged-helix DNA-binding domain, but no significant homologies to any other proteins outside this domain. To identify regions of FAST-1 necessary for its incorporation into ARF, we expressed epitope-tagged deletion mutants of FAST-1 in early embryos by mRNA microinjection and tested them for incorporation into ARF (Fig. 3). In this assay, the winged-helix domain is necessary for DNA binding by the ARF complex; deletions N-terminal to the forkhead domain did not alter incorporation into ARF (not shown), but deletions of the C terminus identify a 127-amino-acid region that is unnecessary for DNA binding of FAST-1 but is necessary for ARF incorporation (Fig. 3).

To compare the domains of FAST-1 that are responsible for its co-precipitation with Smad2 and Smad4 with those needed for ARF incorporation, mRNAs encoding epitope-tagged deletion mutants of FAST-1 were expressed with tagged Smad2 or Smad4 in the presence or absence of activin stimulation (Fig. 3). Deletion of the N-terminal two-thirds of the protein (up to amino acid 365), including the entire winged-helix domain, did not reduce the ligand-dependent association of FAST-1 to Smad2 and Smad4. The domain responsible for this association in a co-precipitation assay localizes to the 127-amino-acid C-terminal domain (380–506) that is necessary for incorporation of FAST-1 into ARF; we refer to this region as the SMAD-interaction domain (SID) of FAST-1. Additional deletions (Δ 208–453, Δ 506–518, Δ 473–518) narrow the region necessary for ligand-dependent association of FAST-1 and Smad2 to amino acids 453–506. Further deletion of this region (Δ 366–473) from the N-terminal side indicate that amino acids C-terminal to position 473 are sufficient for reduced, but still sig-

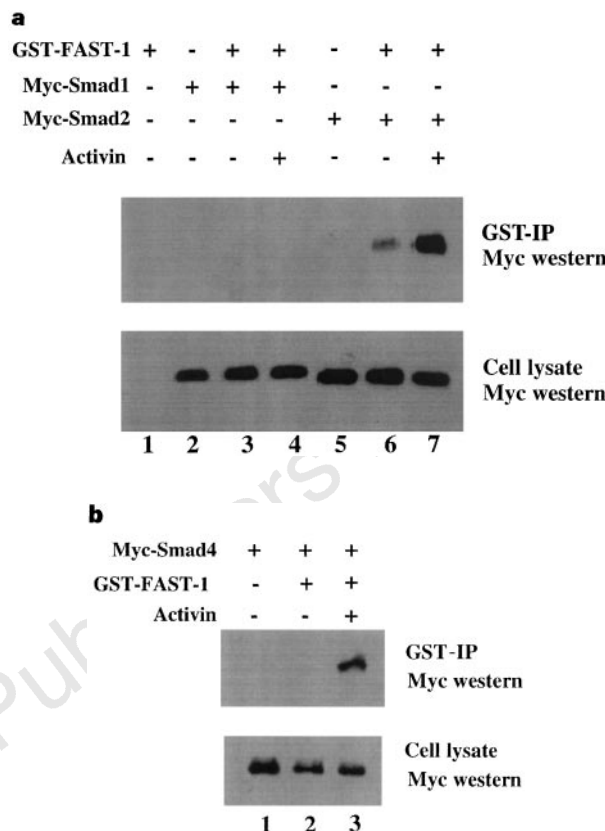


Figure 2 Co-precipitation of SMADs with FAST-1. GST-tagged FAST-1 and Myc-tagged Smad2 or Smad4 were coexpressed in embryos and tested for association. **a**, Co-precipitation of Smad 2 and FAST-1. Top, GST-FAST-1 and Myc-Smad1 or Smad2 were coexpressed with or without activin, immunoprecipitated with anti-GST, and blotted with anti-Myc antibody. Bottom, a portion of the cell lysate was blotted with anti-Myc directly (without immunoprecipitation) to confirm equal expression of Myc-tagged SMADs. **b**, Co-precipitation of Smad4 and FAST-1: GST-FAST-1 and Myc-tagged Smad4 were tested for co-immunoprecipitation as in **a**.

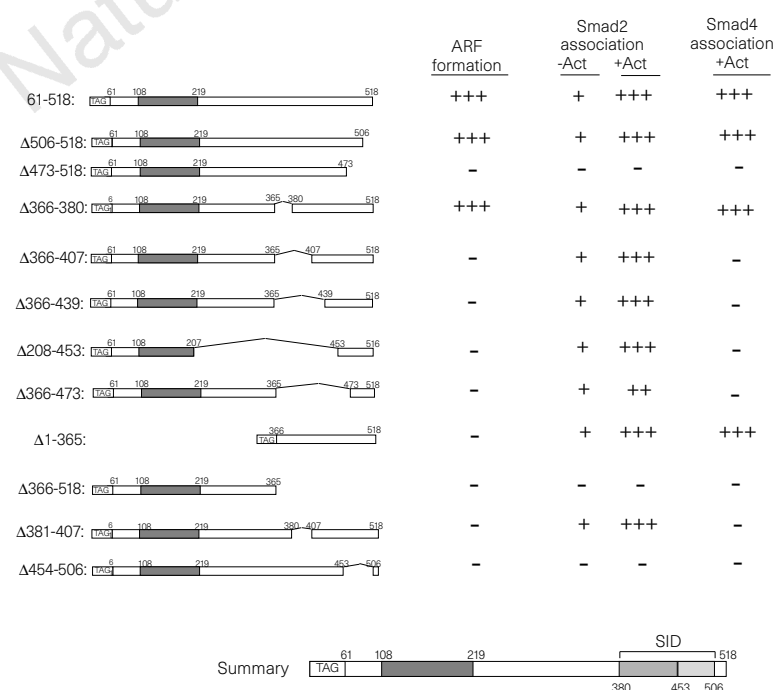


Figure 3 Deletion analysis of FAST-1. FAST-1 incorporation into ARF (in the presence of activin) was tested by EMSA supershift using epitope-tagged FAST-1. Association of FAST-1 with Smad2 and Smad4 was tested by co-immunoprecipitation as for Fig. 2. Shaded boxes denote the winged-helix DNA-binding domain. All constructs generated comparable ARE-binding activity in EMSA, except construct Δ 1–366, which lacked the winged-helix domain. A summary of the deletion results is shown, with the region in FAST-1 necessary for ARF incorporation indicated (SID); this region is necessary for Smad4 association but not for ligand-dependent Smad2 association (left striped box) and a region necessary and sufficient for ligand-dependent Smad2 association (right striped box).

nificant, ligand-dependent association of FAST-1 with Smad2. Comparison of the data for ARF incorporation and co-immunoprecipitation reveals a region (amino acids 380–453) of FAST-1 that is necessary for ARF incorporation but not for ligand-stimulated co-precipitation of Smad2. This region is, however, necessary for the immunoprecipitation of Smad4 with FAST-1; there are no deletion mutants of FAST-1 that immunoprecipitate with Smad4 but not Smad2. These observations indicate that Smad2 and FAST-1 have a ligand-stimulated affinity for one another in the absence of Smad4 in the complex. Phosphorylation of the Smad2 C terminus

may enhance association of the MH2 domain of Smad2 with the 453–506 region of FAST-1 independently of Smad4. Under these circumstances, ARE binding by the complex is still not activated or stable; and additional interaction between Smad4, Smad2 and a region of FAST-1 encompassing amino acids 380–453 is then necessary to stabilize the active (DNA-binding) ARF complex.

The ligand-stimulated co-precipitation of FAST-1, Smad2 and Smad4 confirms that these polypeptides form a complex in the absence of the ARE DNA target, but does not indicate whether additional components of the activin signalling pathway are necessary for complex formation. We tested the affinity of Smad2, Smad4 and FAST-1 for one another in a yeast interaction trap system¹², in which the C-terminal third of FAST-1, where the Smad2 co-immunoprecipitation function of FAST-1 maps (FAST-1 366–518), interacted strongly with the MH2 domain of Smad2 but the winged-helix domain region (FAST-1 56–365) did not (Table 1). Smad1-MH2 and Smad4 itself (in pGAD424, the activator-domain construct) did interact with Smad4 when the latter was expressed in pGBT9 (DNA-binding-domain construct) (Table 1), confirming expression of the activator domain–Smad1 and –Smad4 fusion proteins. The C terminus of FAST-1 did not interact with the MH2 domains of Smad1 or Smad4, confirming the specificity of the interaction with the Smad2 MH2 domain. Additional deletions distinguishing regions necessary for ARF incorporation and Smad4 association from those needed for Smad2 co-immunoprecipitation (such as FAST-1 Δ 366–407) indicate that the region responsible for FAST-1 interaction with Smad2 in yeast is similar to that needed for co-precipitation with Smad2 (Table 1). Although this assay does not exclude the possibility that other proteins may be involved in the Smad2–FAST-1 interaction, as yeast lacks any proteins homologous to components of the activin signalling pathway, our results indicate that additional components of this pathway are not required for Smad2–FAST-1 interaction.

The identification of a domain in FAST-1 that mediates interaction with SMADs raises the possibility of using this domain competitively to inhibit SMAD signalling. mRNA encoding the C-terminal third of FAST-1, containing the SID region, was injected

Table 1 Interaction of FAST-1 and SMADs in yeast		
Bait construct	Interactor construct	Colour intensity (filters)
FAST-1 (aa 61–518)	Smad2 (MH2)	+
	Smad1 (MH2)	–
	Smad4 (FL)	–
	Smad4 (MH2)	–
FAST-1 (aa 366–518)	Smad2 (MH2)	+
	Smad1 (MH2)	–
	Smad4 (FL)	–
	Smad4 (MH2)	–
FAST-1 (aa 56–365: Δ 366–518)	Smad2 (MH2)	–
	Smad1 (MH2)	–
FAST-1 (aa 61–515: Δ 366–407)	Smad2 (MH2)	+
	Smad1 (MH2)	–
FAST-1 (aa 61–515: Δ 366–439)	Smad2 (MH2)	+
	Smad1 (MH2)	–
FAST-1 (aa 61–515: Δ 208–453)	Smad2 (MH2)	+
	Smad1 (MH2)	–
Smad4 (FL)	Smad2 (MH2)	+
	Smad1 (MH2)	+
	Smad4 (FL)	+
Smad2 (MH2)	Smad4 (MH2)	+

Portions of FAST-1 or Smad4 cloned into a Gal4 DNA-binding domain fusion vector were tested for their ability to interact with various Smad–Gal4 activator-domain fusions. Positive results were measured as the development of blue colour on X-gal filter lifts of colonies expressing both activator and DNA-binding domain constructs relative to colonies expressing each construct alone. Filter results were confirmed by quantitative liquid culture assay (not shown). aa, Amino acid residue.

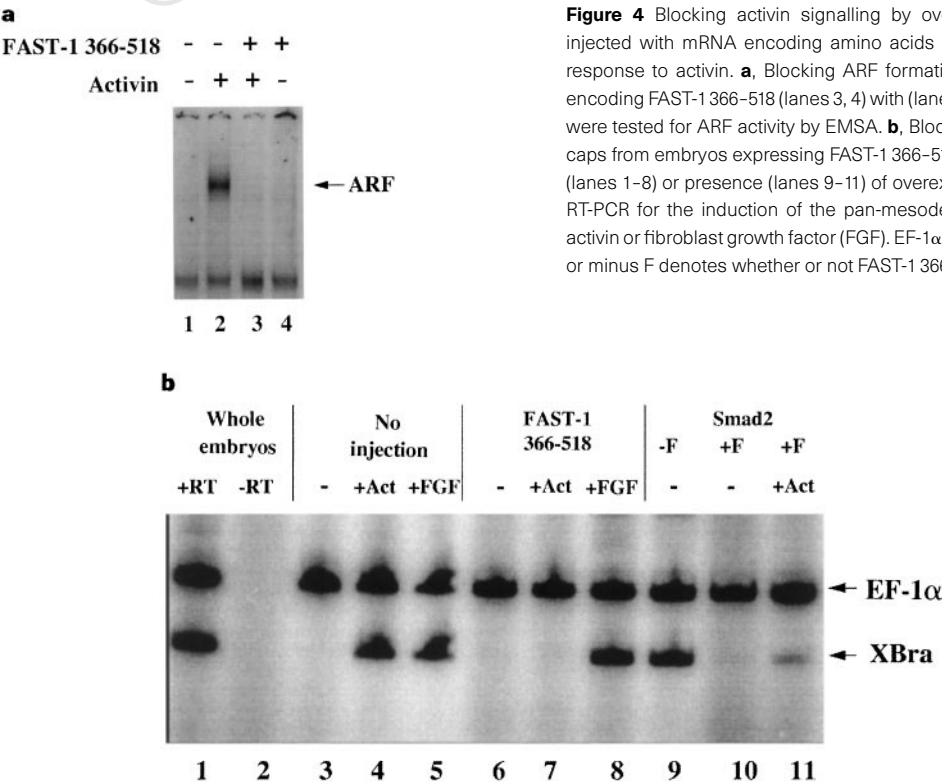


Figure 4 Blocking activin signalling by overexpressed SID. Embryos were injected with mRNA encoding amino acids 366–518 of FAST-1 and tested for response to activin. **a**, Blocking ARF formation. Embryos injected with mRNA encoding FAST-1 366–518 (lanes 3, 4) with (lanes 2, 3) or without (lanes 1, 4) activin were tested for ARF activity by EMSA. **b**, Blocking *brachyury* induction. Animals caps from embryos expressing FAST-1 366–518 (lanes 6–8, 10, 11) in the absence (lanes 1–5) or presence (lanes 9–11) of overexpressed Smad2 were assayed by RT-PCR for the induction of the pan-mesodermal marker *brachyury* (*Xbra*) by activin or fibroblast growth factor (FGF). EF-1α is shown as a loading control. Plus or minus F denotes whether or not FAST-1 366–518 was coinjected with Smad 2.

into the animal pole of 2-cell embryos and tested for its ability to inhibit the formation of ARF and the expression of the early mesodermal marker *brachyury* in response to activin stimulation (Fig. 4a, b). These responses were inhibited by expression of the FAST-1 SID region, as was the elongation of animal caps characteristic of mesoderm induction (not shown). Induction of *brachyury* and of animal cap elongation by fibroblast growth factor was unaffected by the FAST-1 SID (Fig. 4b, and results not shown), indicating that inhibition was specific for the activin/TGF- β signalling pathway. Induction of mesodermal markers by Smad2 overexpression was also inhibited by the FAST-1 SID, but overexpression of Smad2 was sufficient partially to relieve inhibition of activin-stimulated gene expression by the SID (Fig. 4b), in agreement with our interpretation that inhibition is due to the sequestration of Smad2 by the FAST-1 SID.

Smad4, Smad2 and FAST-1 assemble to form a complex that is detectable by gel-shift assay as ARF. The MH2 domain is a sufficient component of Smad2 for ARF formation, but the association of N-terminally truncated Smad2 is activin-stimulated, indicating that the activin regulation of complex formation involves more than simply the relief of inhibition by the MH1 domain. The interaction between the MH2 domain of Smad2 and FAST-1 can be reproduced in yeast, suggesting that it may be direct, whereas interaction between Smad4 and FAST-1 cannot. Smad4 may not interact directly with FAST-1, but rather bind through Smad2, or a conformational change associated with Smad2 binding or some other regulatory modification may be necessary. Mutations in FAST-1 that prevent Smad4 binding do not prevent ligand regulation of Smad2 association with FAST-1 *in vitro*, indicating that activated Smad2 has an affinity for a downstream transcriptional effector even when not associated with Smad4. *In vivo*, however, compartmentalization of FAST-1 and Smad2 may prevent their assembly into a signalling complex until nuclear translocation of Smad2/Smad4 has been stimulated by ligand; Smad4 then seems to be necessary for the formation of ARF as a stable DNA-binding complex. Whether this reflects stabilization of the protein components of the complex or an increase in DNA binding by the complex is unclear.

The affinity of the FAST-1 C terminus for ligand-activated Smad2/Smad4 in intact cells is high enough for it to inhibit activin-TGF- β signalling. Identifying the domains of this region needed for Smad4 and Smad2 binding will help in the design of peptide inhibitors targeting the Smad2 and the Smad4 components of activin/TGF- β signalling. □

Methods

EMSA and co-immunoprecipitation. cDNAs and constructs: HA-tagged Smad4 was provided by A. Hata and J. Massagué; Flag-tagged Smad1, Smad2 (full-length (FL) and MH2 domain) and Smad4 constructs were from P. Hoodless and J. Wrana; Myc-tagged Smad2 was from G. Thomsen; Smad3 was from R. Derynck. *Xenopus* Smad2 (from J. Graff and D. Melton) was fused to 6-Myc tags at amino acid 1 of Smad2 by cloning into the fusion vector pCS2(+)/MT¹³. FAST-1 was tagged by N-terminal fusion at amino acid 61 with either glutathione S-transferase (GST) or Myc tags (pCS2(+)/MT), and Myc-tagged Smad4 was constructed by fusion of full-length Smad4 into pCS2(+)/MT. Sequence information provided by M. Kawabata and confirmed in our laboratory indicates that there is a stop codon in the FAST-1 sequence which terminates the coding sequence at amino acid 518. A corrected sequence has been submitted to Genbank (accession number U70980).

EMSA and supershift assays were performed as described^{10,11}. Embryos were injected at the 2-cell stage with 0.5–2 ng RNA encoding activin and/or tagged SMADs, activin or FAST-1. Embryos were collected for EMSA or co-immunoprecipitation at stage 9 in lysate buffer¹⁰ and cleared by centrifugation for 15 min at 32,000g. Cleared lysates were immunoprecipitated with anti-epitope tag antibody for 1 h at 4 °C, then incubated with protein A–Sepharose for 30 min. Beads were then washed as follows: 1 × lysate buffer plus 0.2 M NaCl and 0.25% N-P40, and 1 × lysate buffer plus 0.4 M NaCl, 1 × lysate buffer plus

0.5% N-P40, 1 × lysate buffer plus 0.2 M NaCl and 0.25% N-P40, and 1 × lysate buffer. Samples were then electrophoresed, transferred to nitrocellulose and detected by ECL as described¹⁴.

Interaction trap constructs. Truncated derivatives of FAST-1 and SMAD genes were cloned into the shuttle/expression vectors pGBT9 and pGAD424 or pGAD10, respectively¹⁵. A fusion of the Gal4 DNA-binding domain in pGBT9 vector with each FAST-1 protein derivative was generated, as was a fusion of the Gal4 activation domain (AD) in the pGAD424/pGAD10 vector with each SMAD.

Specifically, FAST-1–GAL4 DNA-binding domain fusion proteins in pGBT9 included the following regions of FAST-1: (1) FAST-1 N domain and C domain (amino acids 61–518); (2) FAST-1 C terminus (amino acids 366–518); (3) FAST-1 Δ 208–453 (amino acids 61–515, with amino acids 208–453 deleted); (4) FAST-1 Δ 366–407 (amino acids 61–515, with 366–407 deleted); (5) FAST-1 Δ 366–439 (amino acids 61–515, with 366–439 deleted); (6) FAST-1 forkhead domain (amino acids 56–365). SMAD–Gal4 activation-domain fusion proteins in pGAD424 or pGAD10 were generated which included the following SMAD regions: (1) *Xenopus* Smad2 MH2 domain (amino acids 248–467); (2) human Smad1 MH2 domain (amino acids 249–465). Full-length mouse Smad4 and its MH2 domain were additionally cloned into the pGBT9 vector.

Transformation and testing of yeast with two-hybrid clones. Yeast transformations and colony lift filter assays were carried out according to the Matchmaker two-hybrid system protocol (Clontech). For filter assay, colony colour was periodically observed during a 5–6-h incubation at 30 °C following initial exposure of permeabilized yeast to the Z buffer/X-gal solution. Liquid culture β -galactosidase was assayed according to the Matchmaker two-hybrid system protocol.

Animal cap assays and RT-PCR. Embryos were microinjected with 2 ng FAST-1 and/or 150 pg (Smad2) RNA into both blastomeres of 2-cell embryos. Animal caps were cut from stage 8–9 blastulae and cultured in 0.7 × MMR containing 0.1% gelatin, 100 μ g ml⁻¹ BSA, 250 μ g ml⁻¹ Gentamicin (Gibco BRL), and 200 pM purified recombinant activin (Ajinomoto, Inc.) or 100 ng ml⁻¹ human recombinant bFGF (Gibco BRL), until control embryos reached stage 10.5 (ref. 16). Total RNA was extracted at stage 10.5, and RT-PCR was performed as described¹⁷ by using 20 amplification cycles for *EF-1 α* and 25 for *Xbra*.

Received 10 March; accepted 27 June 1997.

- Massagué, J. The TGF β family of receptors. *Cell* **69**, 1067–1070 (1992).
- Macias-Silva, M. et al. Mad2 is a substrate of the TGF- β receptor and its phosphorylation is required for nuclear accumulation and signaling. *Cell* **87**, 1215–1224 (1996).
- Sekelsky, J. J., Newfield, S. J., Raftery, L. A., Chartoff, E. H. & Gelbart, W. M. Genetic characterization and cloning of *Mothers against dpp*, a gene required for decapentaplegic function in *Drosophila melanogaster*. *Genetics* **139**, 1347–1358 (1995).
- Savage, C. et al. *Caenorhabditis elegans* genes *Sma-2*, *Sma-3*, and *Sma-4* define a conserved family of transforming growth factor beta pathway components. *Proc. Natl Acad. Sci. USA* **93**, 790–794 (1996).
- Derynck, R. & Zhang, Y. Intracellular signalling—the Mad way to do it. *Curr. Biol.* **6**, 1226–1229 (1996).
- Massagué, J. TGF- β signaling—receptors, transducers, and Mad proteins. *Cell* **85**, 947–950 (1996).
- Wrana, J. L. & Attisano, L. A. Mad proteins in TGF β signalling. *Trends Genet.* **12**, 493–496 (1996).
- Lagna, G., Hata, A., Hemmati-Brivanlou, A. & Massagué, J. Partnership between Dpc4 and Smad proteins in TGF- β signalling pathways. *Nature* **383**, 832–836 (1996).
- Zhang, Y., Feng, X. H., Wu, R. Y. & Derynck, R. Receptor-associated Mad homologues synergize as effectors of the TGF- β response. *Nature* **383**, 168–172 (1996).
- Huang, H. C., Murtaugh, L. C., Vize, P. D. & Whitman, M. Identification of a potential regulator of early transcriptional responses to mesoderm inducers in the frog embryo. *EMBO J.* **14**, 5965–5973 (1995).
- Chen, X., Rubock, M. J. & Whitman, M. A. Transcriptional partner for Mad proteins in TGF- β signalling. *Nature* **383**, 691–696 (1996).
- Fields, S. & Song, O. A novel genetic system to detect protein–protein interactions. *Nature* **340**, 245–246 (1989).
- Turner, D. L. & Weintraub, H. Expression of *achaete-scute* homolog 3 in *Xenopus* embryos converts ectodermal cells to a neural fate. *Genes Dev.* **8**, 1434–1447 (1994).
- LaBonne, C., Burke, B. & Whitman, M. Role of MAP kinase in mesoderm induction and axial patterning during *Xenopus* development. *Development* **121**, 1475–1486 (1995).
- Bartel, P., Chien, C., Sternglanz, R. & Fields, S. In *Cellular Interactions in Development: A Practical Approach* (ed. Hartley, D.) 153–179 (Oxford University Press, Oxford, 1993).
- Nieuwkoop, P. D. & Faber, J. *Normal Table of Xenopus laevis (Daudin)* (North Holland, Amsterdam, 1967).
- LaBonne, C. & Whitman, M. Mesoderm induction by activin requires FGF mediated intracellular signals. *Development* **120**, 463–472 (1994).

Acknowledgements. We thank M. Kawabata for permission to cite his correction of the FAST sequence, the laboratories of J. Massagué, J. Wrana, R. Derynck and J. Thomsen for cDNAs, M. Rubock for construction of GST-tagged FAST-1, and R. Miale-Lyer for critically reading the manuscript. V. F. was supported by a postdoctoral fellowship from the American Heart Association. M.W. was supported by the Lucille P. Markey Charitable Trust and grants from the NIH.

Correspondence and requests for materials should be addressed to M.W. (e-mail: mwhitman@warren.med.harvard.edu).

Actin-dependent localization of an RNA encoding a cell-fate determinant in yeast

Peter A. Takizawa^{†‡}, Anita Sil[‡], Jason R. Swedlow[†], Ira Herskowitz[‡] & Ronald D. Vale^{†‡}

^{*} Howard Hughes Medical Institute, and Departments of [†] Cellular and Molecular Pharmacology and [‡] Biochemistry and Biophysics, University of California, San Francisco, California 94143-0450, USA

The cytoplasmic localization of messenger RNA creates an asymmetric distribution of proteins that specify cell fate during development in multicellular eukaryotes^{1,2}. The protein Ash1 is a cell-fate determinant in budding yeast which localizes preferentially to the presumptive daughter nucleus, where it inhibits mating-type switching^{3,4}. Here we show that Ash1 mRNA is localized to the distal tip of daughter buds in post-anaphase cells. Three-dimensional imaging reveals that Ash1 mRNA is assembled into particles that associate with the cell cortex. To achieve this localization, Ash1 mRNA must have its 3' untranslated region and the actin cytoskeleton must be intact. Ash1 mRNA is not localized correctly in the absence of a myosin (Myo4) and is mislocalized to the mother-bud neck in the absence of a regulator of the actin cytoskeleton known as Bni1. We propose that Ash1 mRNA particles are transported into the daughter bud along actin filaments and are anchored at the distal tip. Thus, as in higher eukaryotes, *Saccharomyces cerevisiae* employs RNA localization to generate an asymmetric distribution of proteins and hence to determine cell fate.

Transcription of the *ASH1* gene and translation of its RNA occur in large-budded cells that have undergone nuclear division but not cytokinesis^{3,4}. In these post-anaphase cells, Ash1 protein localizes

preferentially to the nucleus in the bud (incipient daughter cell). In principle, asymmetric localization of Ash1 protein could result from restriction of *ASH1* transcription to the incipient daughter nucleus, accumulation of Ash1 mRNA in the incipient daughter cell, or accumulation of Ash1 protein in the incipient daughter as a result of a localization determinant on the protein itself.

To determine whether Ash1 mRNA accumulates in the incipient daughter cell, we looked at the subcellular location of Ash1 mRNA by *in situ* analysis. Standard FISH (fluorescent *in situ* hybridization) techniques were inadequate to detect endogenous Ash1 mRNA. However, by combining a hybridization step employing digoxigenin-labelled antisense RNA probes with an antibody sandwiching procedure, the fluorescent signal was increased sufficiently so that endogenous Ash1 mRNA (Fig. 1), as well as several other RNAs tested (data not shown), could be easily imaged by fluorescence microscopy. Cells deleted for the *ASH1* gene showed only background staining, indicating that the signal was specific (Fig. 1). In an asynchronous population, only the large-budded cells with two distinct DAPI-stained nuclei (indicative of post-anaphase cells) displayed Ash1 mRNA staining (Fig. 1). Synchronization of cells by release from α -factor arrest revealed that the Ash1 mRNA signal peaked in post-anaphase cells and then rapidly declined (data not shown). The presence of Ash1 mRNA in only these cells is consistent with previous results from northern blotting analysis indicating that Ash1 mRNA accumulates in late anaphase and disappears soon afterwards³. Most significantly, the enhanced FISH staining pattern revealed that Ash1 mRNA was concentrated at the distal tip of large buds (Fig. 1). The fidelity of this localization was high: 91% of cells had tightly localized Ash1 mRNA at the distal tip, with virtually no signal in the mother cell; 7% had delocalized mRNA confined to the bud; and only 3% showed significant staining in the mother cell ($n = 100$).

We also examined cells that overexpress *ASH1* and detected not only staining at the distal tip but also significant amounts of unlocalized Ash1 mRNA in both the mother cell and bud (data not shown). This is consistent with the finding that Ash1 protein is

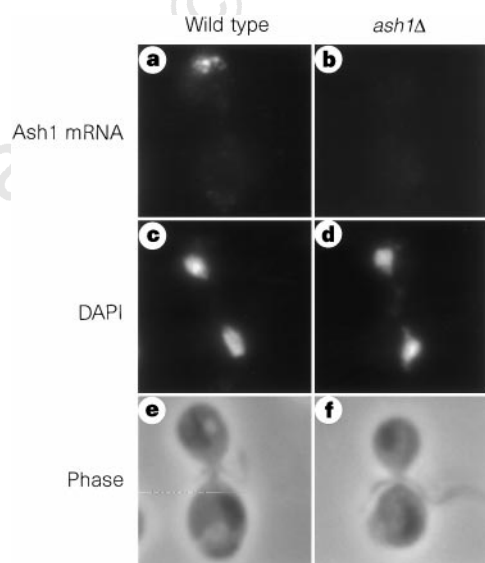


Figure 1 Ash1 mRNA is localized to the distal tip of the presumptive daughter cell. Isogenic wild-type cells (YAS140) (a, c, e) and *ash1Δ* cells (YAS145) (b, d, f) were grown to early log phase, fixed and probed for Ash1 RNA by FISH (a, b). DNA staining by DAPI (4,6-diamino-2-phenylindole) fluorescence (c, d) and phase-contrast images (e, f) indicate that Ash1 mRNA is stained during post-anaphase.

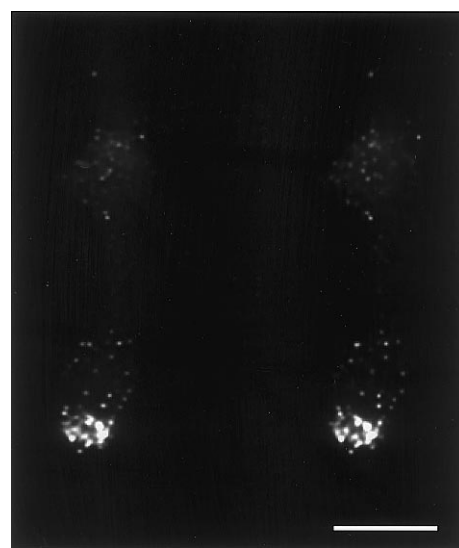


Figure 2 Three-dimensional reconstruction of Ash1 mRNA-containing particles. The image is presented as a stereo pair. To study a uniform population of cells, post-anaphase cells were accumulated by arresting cells with α -factor and then incubated for 60 min in pheromone-free medium. Cells were fixed and processed by FISH to detect Ash1 mRNA. Wide-field optical sectioning and constrained, iterative deconvolution were performed to generate a three-dimensional reconstruction. The background is increased to allow visualization of the mother cell. Scale bar, 5 μ m.

found in both mother and daughter nuclei of cells overexpressing Ash1 (ref. 4). Because overexpression of *ASH1* caused abnormal localization, we visualized Ash1 mRNA from the chromosomal *ASH1* gene in all of the experiments described below.

To study the intracellular distribution of Ash1 mRNA at higher resolution, we imaged FISH staining patterns by wide-field optical sectioning microscopy and three-dimensional image restoration^{5,6}. This imaging technique revealed that Ash1 mRNA was organized into diffraction-limited (<200 nm) particles (Fig. 2), as found for other localized mRNAs⁷⁻⁹; moreover, the particles were clearly localized near the cortex. Even the minority of particles that were not at the distal tip were also located adjacent to the cell membrane (Fig. 2).

To investigate how Ash1 mRNA is transported to and then maintained at the bud tip, we studied Ash1 mRNA localization in mutant strains deleted for the genes *SHE1*–*SHE5*, which are required for proper localization of Ash1 protein^{3,10}. Three of these genes encode proteins that interact with or influence the actin cytoskeleton: *SHE1* encodes a putative actin-based myosin motor, Myo4 protein¹⁰; *SHE5* encodes Bni1 protein, which is involved in promoting actin-filament formation and polarization¹⁰; and *SHE4* encodes an uncharacterized protein necessary for proper polarization of the actin cytoskeleton¹¹. In strains deleted for *MYO4/SHE1*, *SHE2*, *SHE3* or *SHE4*, the Ash1 mRNA particles were distributed throughout the mother and daughter compartments of large-budded cells (Fig. 3, and data not shown). In contrast, in *bni1Δ*/*she5Δ* cells, Ash1 RNA accumulated in the mother-bud neck region of post-anaphase cells (Fig. 3). In a strain lacking both *MYO4/SHE1* and *BNI1/SHE5*, Ash1 mRNA was mislocalized in a way indistinguishable from that observed in the *myo4Δ*/*she1Δ* strain (data not shown). Thus, four of the *she* mutant strains were incapable of localizing Ash1 mRNA, whereas in the *bni1Δ*/*she5Δ* strain, Ash1 mRNA was localized to the neck instead of to the distal tip.

Analysis of the *she* mutants suggested that the actin cytoskeleton might play a role in localizing Ash1 mRNA. To test the requirement for actin filaments in Ash1 mRNA localization, wild-type cells were treated with the actin-depolymerizing agent latrunculin A¹². After a 5-min incubation with 200 μM latrunculin A, the actin cytoskeleton was completely disrupted and Ash1 mRNA was randomly distributed throughout the presumptive mother and daughter compartments (Fig. 4). Latrunculin A also affected the asymmetric accumulation of Ash1 protein; after a 5-min incubation, 70% of larger-budded, post-anaphase cells had equivalent Ash1 protein staining in both mother and daughter nuclei (symmetric staining), 21% had more Ash1 protein in the daughter nucleus than in the mother (partially symmetric), and only 9% showed Ash1 protein

exclusively in the daughter nucleus (asymmetric staining) ($n = 100$). Control cells, in contrast, showed 10% symmetric staining, 19% partially symmetric staining and 71% asymmetric staining ($n = 100$). In cells treated with the microtubule-depolymerizing agent nocodazole (50 μM), microtubules were absent but Ash1 mRNA was still localized to the distal tip of daughter cells (data not shown). Thus, perturbation of the actin cytoskeleton, but not microtubules, disrupts Ash1 mRNA localization and Ash1 protein asymmetry.

In higher eukaryotes, mRNA localization is specified by sequence elements in the 3' untranslated region (UTR)^{1,2}. To see whether the *ASH1* 3'UTR is required for Ash1 mRNA localization, we integrated an allele of *ASH1* lacking its 3'UTR at the *ASH1* locus. Ash1 mRNA expressed from the 3'UTRΔ allele failed to localize to the distal tip; instead it was distributed as particles throughout the mother and daughter (Fig. 3). Furthermore, in 3'UTRΔ cells, the normal

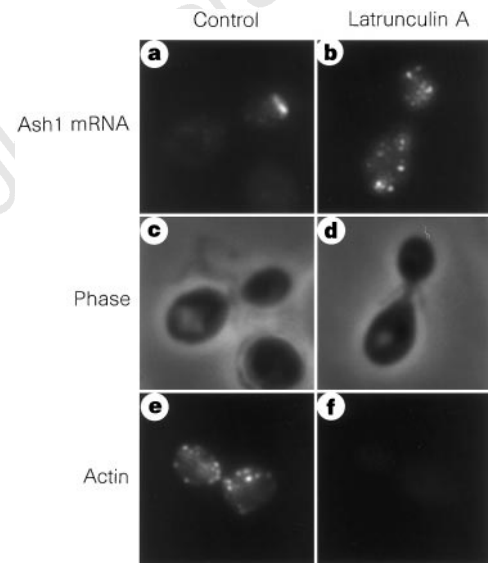


Figure 4 The actin cytoskeleton is required for Ash1 mRNA localization. Cells (YAS140) were grown to early log phase and treated with 200 μM latrunculin A/2% DMSO (b, d, e), or an equivalent amount of DMSO (a, c, f), for 5 min, fixed and probed for Ash1 mRNA by FISH (a, b). An aliquot of each sample was removed after fixation and stained with rhodamine phalloidin¹⁶ to demonstrate that actin filaments are completely depolymerized by this treatment (e, f).

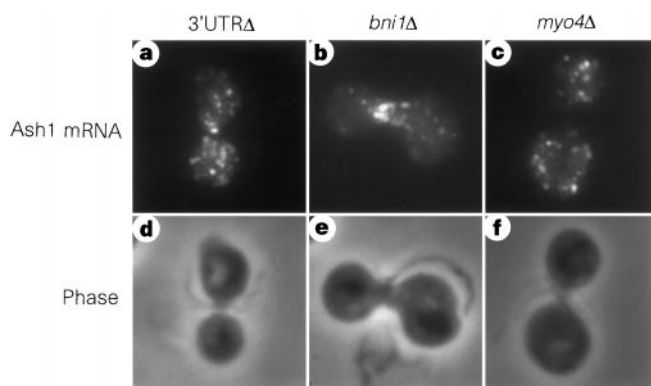


Figure 3 Ash1 mRNA localization requires its 3'UTR and the cytoskeletal proteins, Bni1 and Myo4. 3'UTRΔ (YAS222) (a, d), *she5Δ*/*bni1Δ* (YAS219-1A) (b, e), and *she1Δ*/*myo4Δ* (YAS217-1A) (c, f) were grown to early log phase, fixed and probed for Ash1 mRNA by FISH (a, b, c). The corresponding phase-contrast images are shown (d, e, f).

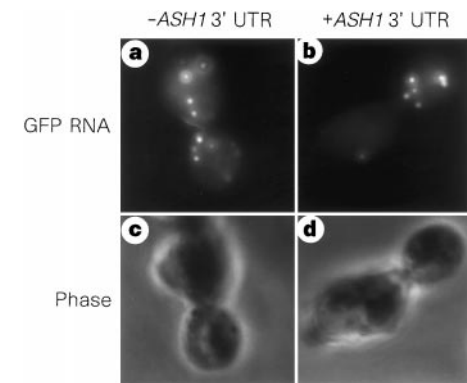


Figure 5 The *ASH1* 3'UTR is sufficient for asymmetric RNA localization to the bud. *ash1Δ* cells (YAS145) carrying either *pPHO4-GFP* (EB0359) (a, c) or *pPHO4-GFP* fused to the *ASH1* 3'UTR (pPT100) (b, d) were synchronized (see Methods) and subjected to FISH analysis (a, b). The corresponding phase images are shown (c, d).

asymmetry of Ash1 protein was markedly disrupted: 80% of large-budded, post-anaphase cells showed symmetric staining, 5% showed partially symmetric staining, and only 15% showed asymmetric staining ($n = 200$). Wild-type cells, in contrast, gave only 4% symmetric cells, 19% partially symmetric cells and 77% exclusively asymmetric cells ($n = 200$). Thus, the 3'UTR is required for proper localization of Ash1 mRNA and protein.

We next investigated whether the *ASH1* 3'UTR was sufficient to localize a heterologous RNA to daughter cells. Cells were transformed with a plasmid encoding either green fluorescent protein (GFP) or GFP fused to a 564-nucleotide fragment containing the *ASH1* 3'UTR. The GFP mRNA was detected by FISH in a synchronized population of large-budded cells. We found that 65% of the cells containing the GFP-*ASH1* 3'UTR fusion showed asymmetric localization of GFP mRNA to the daughter cell ($n = 100$) (Fig. 5). In contrast, GFP mRNA without the *ASH1* 3'UTR was asymmetrically localized in only 6% of cells ($n = 100$) (Fig. 5). Shorter regions of the *ASH1* 3'UTR fused to GFP (the 5' and 3' 282-nucleotide halves of the 3'UTR fragment described above) likewise failed to localize asymmetrically (not shown). Although the *ASH1* 3'UTR was sufficient to confer asymmetric localization of GFP mRNA to the bud, the GFP mRNA failed to concentrate at the distal tip as seen for Ash1 mRNA. Thus, it appears that the *ASH1* 3'UTR can specify transport to daughter cells, but other sequences in Ash1 mRNA may be required to anchor it at the distal tip.

We have shown that Ash1 mRNA is localized to the distal tip of the incipient daughter cell. Under conditions in which Ash1 mRNA localization is disrupted (as a result of overexpression of *ASH1*, deletion of the 3'UTR or treatment with latrunculin A, or in *she* mutant strains), Ash1 protein accumulates symmetrically in both mother and daughter nuclei. We therefore propose that asymmetric localization of the cell-fate determinant Ash1 results from localization of its mRNA. Rapid localization of Ash1 mRNA to the distal pole of the predivisional cell might be sufficient to restrict most Ash1 to the daughter-cell nucleus; once the mRNA is translated, most of the resultant protein will translocate into the most proximal nucleus.

Our results suggest that Ash1 mRNA localization occurs as a result of transport of RNA-containing particles along actin filaments to the distal tip of the bud. We propose that movement of RNA to the distal pole may be very rapid, because synchronized cells do not show a well defined intermediate population of cells with RNA particles in transit to the bud. Once at the distal pole, the RNA might be tethered to the cortex by interactions with an anchor protein. In *bni1Δ* cells, the anchor protein may be trapped at the mother-bud neck, thereby sequestering Ash1 mRNA at an inappropriate site. None of the *She* proteins is localized exclusively to the distal tip at the time Ash1 mRNA accumulates¹⁰, suggesting that the anchor protein is yet to be identified.

The mechanism of mRNA localization in *S. cerevisiae* appears to be similar to that in higher eukaryotes, in which microtubules and actin filaments have been implicated in the transport and anchoring of RNA^{9,13–15}. The ease of genetic and biochemical manipulation of *S. cerevisiae* makes it a propitious system for determining the details of how mRNA is localized. Moreover, our results suggest that other yeast proteins involved in cell-fate determination or bud formation may be asymmetrically distributed by transport and localization of their mRNAs.

Note added in proof: Another report of Ash1 mRNA localization appeared while this Letter was under review¹⁷. □

Methods

Yeast strains. All strains were derived from YAS140 which is W303a (K699: *a ura3-1 his3-11,15 trp1-1 leu2-3,112 ade2-1 can1-100*). YAS145: *a ash1ΔLEU2*. YAS217-1A: *a she1ΔURA3*. YAS218-1D: *a she3ΔURA3*. YAS219-1A: *a she5ΔURA3*. YAS220-1B: *a she2ΔURA3*. YAS221-3A: *a she4ΔURS3*. YAS222: *a ash1ΔLEU2::ASH1-3'UTRΔ-URA3*. (YAS222 contains an allele of

ASH1 that lacks its 3'UTR; this allele has been integrated at the *ASH1* locus in place of the wild-type gene.) The following strains were a kind gift from R. P. Jansen and K. Nasmyth: YAS171 (K5102: *a/α SHE5/she5::URA3*); YAS172 (K5104: *a/α SHE1/she1::URA3*); YAS173 (K5213; *a/α SHE3/she3::URA3*); YAS174 (K5477: *α she2::URA3*); YAS175 (K5560: *α she4::URA3*).

Sample preparation and enhanced FISH. Cultures were grown to OD₆₀₀ ~0.5 and fixed in 4% formaldehyde for 1 h. Cells were washed and spheroplasted in 100 mM potassium phosphate buffer, pH 7.0, 1.2 M sorbitol containing 30 mM β-mercaptoethanol and 40 μg ml⁻¹ zymolyase 100T (ICN Biomedicals) for 15 min at 37 °C. Cells were washed and spread on polylysine-coated, multiwell test slides. Cells were incubated for 1 h in hybridization mix (50% formamide, 5 × SSC, 1 mg ml⁻¹ yeast tRNA, 100 μg ml⁻¹ heparin, 1 × Denhardt's, 0.1% Tween-20, 0.1% CHAPS, and 5 mM EDTA), then incubated overnight at 37 °C in hybridization mix containing 5 μg ml⁻¹ Ash1 probe. Ash1 probe was prepared by generating digoxigenin-labelled, antisense Ash1 RNA with Megascript Kit (Ambion) according to the manufacturer's protocol. The resultant digoxigenin-labelled RNA was subjected to alkaline hydrolysis to generate ~0.25-kb fragments. After hybridization, cells were washed in 0.2 × SSC and blocked in PBS, 0.1% Triton X-100, 10% horse serum. Cells were incubated with mouse anti-digoxigenin antibody (Boehringer-Mannheim) diluted to 0.4 μg ml⁻¹ in blocking buffer for 2 h. After washing in PBS, cells were incubated with rabbit anti-mouse IgG diluted to 1 μg ml⁻¹ in blocking buffer for 1 h. After washing, cells were incubated with rhodamine-conjugated goat anti-rabbit IgG (ICN Pharmaceuticals). Cells were washed and mounted in 0.01 M Tris-Cl, pH 8.4, 90% glycerol, 1 mg ml⁻¹ *p*-phenylenediamine, and 0.1 μg ml⁻¹ DAPI (Sigma). We also used a second detection method that results in a considerably stronger signal: after overnight hybridization and washing, cells were incubated with anti-digoxigenin coupled to alkaline phosphatase (Boehringer-Mannheim). A fluorescent precipitate was generated using an HNPP fluorescent detection kit (Boehringer-Mannheim).

Synchronized cells were generated by growing cells to OD₆₀₀ = 0.3. α-Factor was added to 20 μg ml⁻¹ and cultures were incubated for 3 h at 30 °C. Cells were washed twice in YPD and incubated for 60 min at 30 °C. Cells were fixed and processed as described.

Microscopy. Samples were viewed under a Nikon EFD-3 microscope using a 100×/NA 1.4 lens. Images were captured with a cooled charge-coupled device, and digital images were displayed using Adobe Photoshop. Wide-field optical sectioning and image deconvolution were performed by collecting optical sections with a Nikon 100×/NA 1.4 lens at 0.2-μm intervals on a wide-field microscopy workstation equipped with a Peltier-cooled charge-coupled device (Kodak) mounted in a scientific imaging camera (Photometrics)⁵. Images were restored by constrained, iterative deconvolution⁶. This method uses an empirical measurement of the image blurring caused by the microscope objective lens to restore out-of-focus photons to their correct position. Stereo pairs were calculated by summing pixel values along projection rays⁵.

Received 23 June; accepted 7 August 1997.

1. St Johnston, D. The intracellular localization of messenger RNAs. *Cell* **81**, 161–170 (1995).
2. Wilhelm, J. E. & Vale, R. D. RNA on the move: the mRNA localization pathway. *J. Cell Biol.* **123**, 269–274 (1993).
3. Bobola, N., Jansen, R.-P., Shin, T. H. & Nasmyth, K. Asymmetric accumulation of Ash1p in postanaphase nuclei depends on a myosin and restricts yeast mating-type switching to mother cells. *Cell* **84**, 699–709 (1996).
4. Sil, A. & Herskowitz, I. Identification of an asymmetrically localized determinant, Ash1p, required for lineage-specific transcription of the yeast *HO* gene. *Cell* **84**, 711–722 (1996).
5. Swedlow, J. R., Sedat, J. W. & Agard, D. A. In *Deconvolution of Images and Spectra* 2nd edn 284–309 (Academic Press, New York, 1997).
6. Agard, D. A., Hiraoka, Y., Shaw, P. & Sedat, J. W. *Meth. Cell Biol.* **30**, 353–366 (1989).
7. Ainger, K. et al. Transport and localization of exogenous myelin basic protein mRNA microinjected into oligodendrocytes. *J. Cell Biol.* **123**, 431–441 (1993).
8. Taneja, K. L., Lifshitz, L. M., Fay, F. S. & Singer, R. H. Poly(A) RNA codistribution with microfilaments: evaluation by *in situ* hybridization and quantitative digital imaging microscopy. *J. Cell Biol.* **119**, 1245–1260 (1992).
9. Ferrandon, D., Elphick, L., Nusslein-Volhard, C. & St Johnston, D. Stufen protein associates with the 3'UTR of bicoid mRNA to form particles that move in a microtubule-dependent manner. *Cell* **79**, 1221–1232 (1994).
10. Jansen, R.-P., Dowzer, C., Michaelis, C., Galova, M. & Nasmyth, K. Mother cell-specific *HO* expression in budding yeast depends on the unconventional myosin Myo4p and other cytoplasmic proteins. *Cell* **84**, 687–697 (1996).
11. Wendland, B., McCaffery, J. M., Xiao, Q. & Emr, S. D. A novel fluorescence-activated cell sorter-based screen for yeast endocytosis mutants identifies a yeast homologue of mammalian eps15. *J. Cell Biol.* **135**, 1485–1500 (1996).
12. Ayscough, K. R. et al. High rates of actin filament turnover in budding yeast and roles for actin in establishment and maintenance of cell polarity revealed using the actin inhibitor latrunculin-A. *J. Cell Biol.* **137**, 399–416 (1997).

13. Yisraeli, J. K., Sokol, S. & Melton, D. A. A two-step model for the localization of maternal mRNA in *Xenopus* oocytes: involvement of microtubules and microfilaments in the translocation and anchoring of Vg1 mRNA. *Development* **108**, 289–298 (1990).
14. Sundell, C. L. & Singer, R. H. Requirement of microfilaments in sorting of actin messenger RNA. *Science* **253**, 1275–1277 (1991).
15. Pokrywka, N. J. & Stephenson, E. C. Microtubules are a general component of mRNA localization systems in *Drosophila* oocytes. *Dev. Biol.* **167**, 363–370 (1995).
16. Adams, A. E. M. & Pringle, J. R. in *Guide to Yeast Genetic and Molecular Biology* 729–731 (Academic Press, New York, 1991).
17. Long, R. M. *et al.* Mating type switching in yeast controlled by asymmetric localization of *ASH1* mRNA. *Science* **277**, 383–387 (1997).

Acknowledgements. We thank J. Sedat and D. Agard for use of their CCD-based optical sectioning microscope and image processing facilities; R. P. Jansen and K. Nasmyth for strains; E. O'Shea for plasmids; and D. Drubin for latrunculin A. This work was supported by grants from the NIH (R.D.V. and L.H.) and a grant from the Cancer Research Fund of the Damon Runyon–Walter Winchell Foundation Fellowship (J.R.S.).

Correspondence and requests for materials should be addressed to R.D.V. (e-mail: vale@phy.ucsf.edu).

Reversal in the direction of movement of a molecular motor

Ulrike Henningsen & Manfred Schliwa

Adolf-Butenandt-Institut, Zellbiologie, University of Munich, Schillerstrasse 42, 80336 Munich, Germany

Kinesin and non-claret disjunctional (*ncd*) are molecular motors of the kinesin superfamily that move in opposite directions along microtubules. The molecular basis underlying the direction of movement is unclear, although it is thought to be an intrinsic property of the motor domain, a conserved region about 330 amino acids in length^{1–3}. The motor domain is found at the amino terminus in conventional kinesins, but at the carboxy terminus in *ncd*^{4,5}. Here we report on a chimaera composed of the motor domain of the minus-end-directed kinesin of *Neurospora crassa*⁶. The bacterially expressed fusion protein was tested in motility assays using polarity-marked microtubules⁷. Surprisingly, the chimaera moved towards the plus end, demonstrating that the polarity of force generation of the *ncd* motor domain has been reversed. This finding indicates that the domain organization,

particularly the position of the motor domain, is of fundamental importance for the polarity of force production. It also demonstrates that the direction of microtubule movement is not controlled solely by the motor domain⁸.

The molecular motor non-claret disjunctional (*ncd*) is a member of the kinesin superfamily that moves towards the minus end of microtubules. It consists of an N-terminal tail region of ~200 amino acids, a central stalk with heptad repeats, as in the stalk of conventional kinesin, and a C terminus that contains a region (residues 348–670) that is 41% identical to the motor domain of other kinesins^{5,9}. We constructed a chimaera in the vector pt7-7 (ref. 10) consisting of the motor domain of *ncd* derived from glutathione S-transferase (GST)–MC5 (ref. 11) fused to the N terminus of the stalk/tail domain of *Neurospora* kinesin (Nkin). In this chimaera, the first amino acid is Met 333 of the *ncd* sequence (numbering according to ref. 9), and the splice site is within the codon for Leu 661 in *ncd* (Leu 321 in Nkin). This leucine is located in a conserved region of the helix α_6 , which is identical in length and position in the crystal structures of kinesin and *ncd*¹². Therefore, neither the length nor the characteristics of this helix are expected to change in the chimaera. The juncture is located seven amino acids N-terminal to the end of the conserved motor domain of most kinesins^{1,2,12}. The resulting product, which showed the splice site in the expected position (Fig. 1), was sequenced and shown to be free of error. A set of digests of the chimaera with appropriate restriction enzymes, as well as polymerase chain reactions (PCR) using internal primers, all yielded the expected products (not shown), supporting the conclusion that the plasmid was constructed as we thought.

The *ncd*–Nkin chimaera was expressed in *Escherichia coli*. Induction with IPTG yielded large amounts of a soluble polypeptide with a relative molecular mass of about 105,000 (M_r 105K), which is of the expected size for the *ncd*–Nkin chimaera. This polypeptide could be purified by a microtubule affinity step in the presence of apyrase and the non-hydrolysable ATP analogue AMP-PNP (Fig. 2). The material released from microtubules by ATP (S5) consisted almost exclusively of the 105K polypeptide, with contamination from tubulin. Thus this polypeptide exhibited the ATP-dependent microtubule binding and release behaviour typically of microtubule motors. Its behaviour on a Superose 6 gel-filtration column was very similar to that of bacterially expressed Nkin, and suggests that the

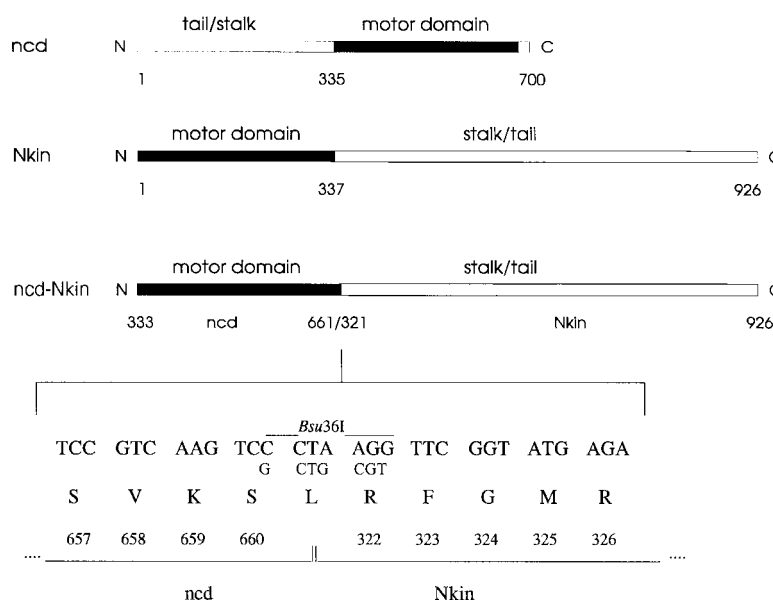


Figure 1 Overview of the domain organization of the non-claret disjunctional motor (*ncd*), *Neurospora* kinesin (Nkin), and the chimaera (*ncd*–Nkin). The nucleotide sequence of the head/stalk transition of *ncd*–Nkin (as determined by sequencing of the construct) is shown at the bottom. The *Bsu361* site generated

by site-directed mutagenesis is indicated, along with the original nucleotide sequences, shown in smaller print. The codon for Leu661/321 is CTG in both *ncd* and Nkin.

molecule was a dimer with a Stokes radius of 7.0 nm (not shown).

The identity of this polypeptide was investigated in immunoblots using a polyclonal antibody against the Nkin stalk/tail domain⁶ and a peptide antibody (MMR48) against a consensus sequence 24 amino acids in length in the HIPYR region of animal kinesins¹³. The Nkin antibody recognized Nkin as well as the chimaera, but not GST-MC5 (ncd). Antibody MMR48 reacted strongly with Nkin, which was to be expected given the high degree of sequence conservation of the HIPYR peptide between animal kinesins and Nkin. Its reaction with GST-MC5, however, was much weaker, because of considerable sequence divergence in the HIPYR peptide between ncd and animal kinesins. The chimaera showed the same weak reaction with MMR48 as GST-MC5. Thus the chimaera combines the immunological characteristics of the Nkin stalk/tail domain and the ncd motor domain.

A standard microtubule gliding assay⁶ was used to determine the motile properties of ncd-Nkin. The data are summarized in Table 1, with the results of gliding assays using the two sources of the chimaera, Nkin and GST-MC5. Both the high-speed bacterial extract (S2) and the microtubule release (S5) of ncd-Nkin produced continuous microtubule gliding that could be followed for

several minutes. Overall, fewer microtubules attached to the coverslip (less than one-third compared to Nkin or GST-MC5) despite the presence of high levels of ncd-Nkin, but microtubules that attached were able to move. As time passed, however, more and more microtubules remained immotile or detached from the coverslip. Microtubule gliding produced by S5 material occurred at a slightly higher average velocity, distributed over a broader range, although microtubule attachment to the coverslip was lower than in S2. The reasons for this are unclear, but it is conceivable that the chimaera is more sensitive to the nucleotide depletion step required for the binding of the motor to microtubules, and so more motor molecules are inactive.

To determine the direction of movement of this construct, motility assays were performed using polarity-marked microtubules⁷ in which the polymer added to the brightly fluorescent microtubule seed is only dimly labelled and much shorter at the microtubule minus end. As expected, Nkin and GST-MC5 moved the asymmetrically labelled fluorescent microtubules with the bright seed leading (plus-end directed) and bright seed trailing (minus-end directed), respectively. A high-speed supernatant of bacteria expressing ncd-Nkin as well as S5 release moved micro-

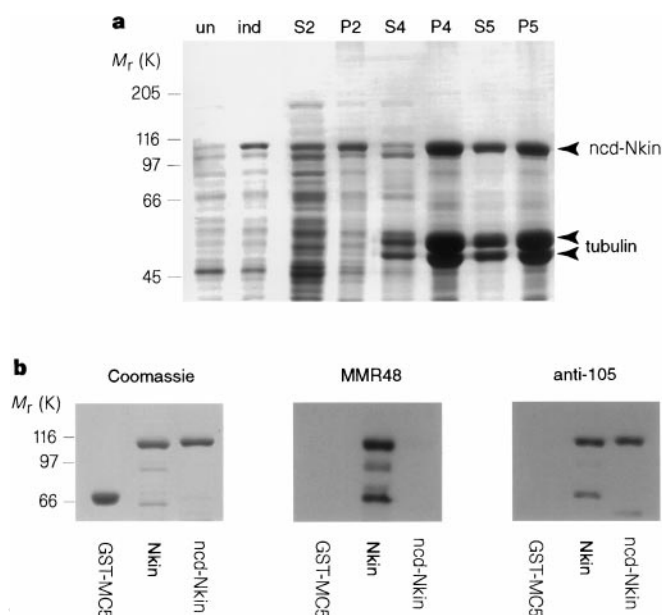


Figure 2 a, Isolation of the ncd-Nkin chimaera expressed in bacteria. The isolation protocol and nomenclature are as described⁶. Columns show uninduced (un) and IPTG-induced (ind) bacteria, and supernatant (s) and pellet (p) of different stages in the isolation protocol. Both the bacterial extract (S2) and the material released from microtubules by ATP (S5) were motile. The positions of ncd-Nkin and tubulin are indicated on the right, and molecular markers on the left. **b**, Immunological characterization of GST-MC5 (ncd), Nkin and the chimaera (ncd-Nkin). Coomassie-stained gel of S5 (left) and corresponding immunoblots reacted with antibody MMR48 against a peptide from the HIPYR-region¹² (middle) and anti-105K antibody against Nkin⁶ (right). MMR48 reacts strongly with Nkin (as well as some degradation products), whereas the reaction with GST-MC5 (ncd) and ncd-Nkin is weaker and is detected only after longer exposure. Anti-105 reacts with Nkin and ncd-Nkin, but not with GST-MC5 (ncd).

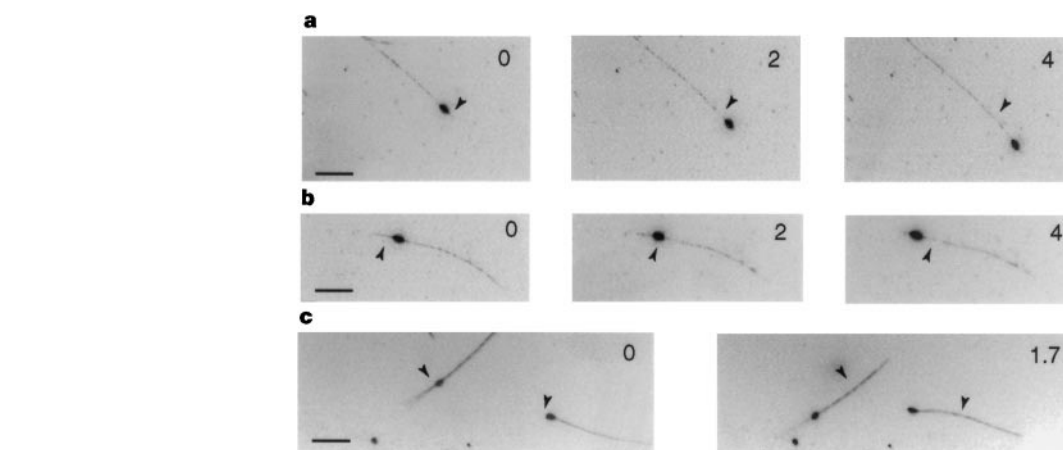


Figure 3 Directionality of ncd-Nkin motor activity using polarity-marked microtubules. **a**, **b**, Gliding produced by S2; **c**, gliding on S5. The images are shown in reverse contrast for better visibility of the weakly labelled microtubule

segments. The asymmetrically labelled microtubules move with the bright seed leading. Time (in min) is indicated in the top right-hand corner. Scale bar 5 μ m.

Table 1 Microtubule gliding velocity and direction of movement

Motor	Velocity ($\mu\text{m min}^{-1}$)	Range	Direction
Nkin	156 $\pm$ 18 (n = 20)	120–186	plus
GST-MC5	5.9 $\pm$ 0.4 (n = 10)	5.3–6.7	minus
ncd-Nkin, bacterial extract	1.7 $\pm$ 0.5 (n = 35)	0.5–2.7	plus
ncd-Nkin, S5	2.7 $\pm$ 1.1 (n = 10)	1.3–4.6	plus

tubules exclusively with the bright seed leading, and this movement could be followed for many minutes (Fig. 3). The velocity of gliding was in the same range as that shown in Table 1. Thus, the head domain of the minus-end-directed motor ncd linked to the N terminus of the stalk of a conventional kinesin produces plus-end-directed movement along a microtubule.

The crystal structures of the kinesin and ncd motor domains are virtually identical^{12,14}, but they differ with respect to two key features: the position of the motor domain in the primary sequence; and the length and nature of several surface loops extending from the core structure of the motor domain. We have shown that the molecular topology of the motor domain relative to the rest of the molecule is, in principle, responsible for the directionality of the motor. Thus the direction of movement is not an exclusive property of the motor domain, although our observations do not rule out the presence of polarity determinants in the motor domain. Kinetic and structural studies showed the enzymatic properties of ncd and kinesin to be very similar with respect to ATPase kinetics and mode of coupling to microtubule binding^{15–17}, suggesting that the opposite polarity for force generation has a structural rather than an enzymatic basis. Consistent with this suggestion, cryo-electron microscopic studies of microtubules with tightly bound dimeric kinesin or ncd demonstrated a striking difference in the position of the unbound head^{18–20}: it points towards the plus end in kinesin but is tilted towards the minus end in ncd. This suggests that the juncture of the head to the rest of the molecule (N or C terminal) influences the topology of microtubule binding of the two-headed construct and, therefore, presumably also the direction of its movement. However, the mode of linkage to the stalk may not be the only factor involved. Our experiments suggest that the chimaera is a relatively poor motor, being slower than ncd and tending to stall and detach more readily. Therefore other features, such as the surface loops, may be important for the fine tuning of motor activity.

The reverse experiment (placing the ncd stalk at the N terminus of a conventional kinesin motor domain) yielded a construct that still moved towards the microtubule plus end (S. A. Endow and G. Rai, personal communication). Similarly, N-terminal fusions of spectrin or GST with the kinesin motor domain resulted in either immobile or plus-end-directed chimaeras⁸. It is possible that these 'reverse' chimaeras have general mechanistic problems and are unable to generate force in the same way as the chimaera we have described. Alternatively, a different splice junction may be more successful. It is also conceivable that sequences outside the motor domain have some effect on motor activity. C-terminal truncation experiments have shown that the efficiency of microtubule movement declines dramatically if the truncation site is close to the conserved motor domain^{8,21,22}. This suggests that sequence elements of the stalk close to the motor domain support motor activity in as yet unknown ways. □

Methods

Cloning. Full-length Nkin cDNA was cloned into *Bam*HI-digested and blunted pT7-7 vector¹⁰. A *Bsu*36I restriction site was created at position 957–966

without changing the corresponding amino-acid sequence (see Fig. 1). The motor domain of ncd (333–660) was amplified by PCR using GST-MC5 (ref. 11) as a template, and primers that introduced an *Nde*I site to the 5'-end and a *Bsu*36I site to the 3'-end of the resulting product. The motor domain of Nkin (amino acids 1–320) was cut out using the *Nde*I site in the polylinker and the newly created *Bsu*36I site, and the ncd motor domain (amino acids 333–660) was inserted instead. The correctness of the ncd–Nkin plasmid was confirmed by sequencing, additional PCR reactions using internal primers, and restriction digestion patterns of the plasmid using several restriction enzymes.

Expression. *E. coli* BL21 cells were grown at 37 °C and induced by adding 0.5 mM IPTG overnight at 21 °C. Cells were pelleted and resuspended in 4 vols per g wet weight of AP100 (ref. 6). Cells were lysed by sonication and centrifuged at 150,000g for 60 min at 4 °C, and the supernatant was collected. Microtubules prepared from phosphocellulose-purified porcine brain tubulin⁶ were stabilized with taxol and added at a concentration of 0.5 mg ml⁻¹. Additional taxol (10 μ M) as well as apyrase (5 U ml⁻¹) and AMP-PNP (2 mM) were added to the supernatant, and the mixture was incubated on ice for 15 min. The microtubules were sedimented (100,000g, 30 min), resuspended in AP100/50 mM KCl, and centrifuged at 100,000g for 30 min through a sucrose cushion. The microtubule pellet (P4) was resuspended in AP100 with 10 mM MgATP. After centrifugation (100,000g, 30 min) the resulting ATP release (S5) contained the ncd–Nkin chimaera. Western blotting was performed as described⁶.

Motility assays. Brightly labelled microtubule seeds were created by assembling rhodamine-labelled tubulin in AP100 with 1 mM GTP, 4 mM MgCl₂ and 40% glycerol at 37 °C for 5 min. The labelled seeds were elongated by dilution into a 1 mg ml⁻¹ unlabelled mixture of NEM-treated tubulin and untreated tubulin (0.7:1.0) prewarmed to 37 °C. After 10 min the labelled microtubules were stabilized by the addition of three vols AP100 plus 20 μ M taxol. Motility assays were performed as described⁶. When using rhodamine-labelled microtubules, 0.1 mg ml⁻¹ catalase, 0.03 mg ml⁻¹ glucose oxidase, 10 mM glucose and 0.3% 2-mercaptoethanol were added to prevent microtubule photobleaching and breaking.

Received 11 June; accepted 16 July 1997.

- Bloom, G. & Endow, S. Motor proteins 1: kinesin. *Prot. Profiles* **1**, 1059–1116 (1994).
- Goldstein, L. S. B. With apologies to Scheherazade: Tails of 1001 kinesin motors. *Annu. Rev. Genet.* **27**, 319–351 (1993).
- Moore, J. D. & Endow, S. A. Kinesin proteins: a phylum of motors for microtubule-based motility. *BioEssays* **18**, 207–219 (1996).
- Walker, R. A., Salmon, E. D. & Endow, S. A. The *Drosophila* claret segregation protein is a minus-end directed motor molecule. *Nature* **347**, 780–782 (1990).
- McDonald, H. B., Stewart, R. J. & Goldstein, L. S. B. The kinesin-like ncd protein of *Drosophila* is a minus end-directed microtubule motor. *Cell* **63**, 1159–1165 (1990).
- Steinberg, G. & Schliwa, M. The *Neurospora* organelle motor: A distant relative of conventional kinesin with unconventional properties. *Mol. Biol. Cell* **6**, 1605–1618 (1995).
- Hyman, A. A. Preparation of marked microtubules for the assay of the polarity of microtubule-based motors by fluorescence. *J. Cell Sci. (suppl.)* **14**, 125–127 (1991).
- Stewart, R. J., Thaler, J. P. & Goldstein, L. S. B. Direction of microtubule movement is an intrinsic property of the motor domains of kinesin heavy chain and *Drosophila* ncd protein. *Proc. Natl Acad. Sci. USA* **90**, 5209–5213 (1993).
- Endow, S. A., Henikoff, S. & Soler-Niedziela, L. Mediation of meiotic and early mitotic chromosome segregation in *Drosophila* by a protein related to kinesin. *Nature* **345**, 81–83 (1990).
- Tabor, S. in *Current Protocols in Molecular Biology* (eds Ausubel, F. A. et al.) 16.2.1–16.2.11 (Green/Wiley, New York, 1990).
- Chandra, R., Salmon, E. D., Erickson, H. P., Lockhart, A. & Endow, S. A. Structural and functional domains of the *Drosophila* ncd microtubule motor protein. *J. Biol. Chem.* **268**, 9005–9013 (1993).
- Sablin, E. P., Kull, F. J., Cooke, R., Vale, R. D. & Fletterick, R. J. Crystal structure of the motor domain of the kinesin-related motor ncd. *Nature* **380**, 555–559 (1996).
- Marks, D. L., Larkin, J. M. & McNiven, M. A. Association of kinesin with the Golgi apparatus in rat hepatocytes. *J. Cell Sci.* **107**, 2417–2426 (1994).
- Kull, J., Sablin, E. P., Lau, R., Fletterick, R. J. & Vale, R. D. Crystal structure of the kinesin motor domain reveals a structural similarity to myosin. *Nature* **380**, 550–555 (1996).
- Crevel, I. M.-T. C., Lockhart, A. & Cross, R. A. Weak and strong states of kinesin and ncd. *J. Mol. Biol.* **257**, 66–76 (1996).
- Lockhart, A. & Cross, R. A. Origins of reversed directionality in the ncd molecular motor. *EMBO J.* **13**, 751–757 (1994).
- Lockhart, A., Crevel, I. M.-T. C. & Cross, R. A. Kinesin and ncd bind through a single head to microtubules and compete for a shared MT binding site. *J. Mol. Biol.* **249**, 763–771 (1995).
- Arnal, I., Metoz, F., DeBonis, S. & Wade, R. H. Three-dimensional structure of functional motor proteins on microtubules. *Curr. Biol.* **6**, 1265–1270 (1996).
- Hirose, K., Lockhart, A., Cross, R. A. & Amos, L. A. Three-dimensional cryoelectron microscopy of dimeric kinesin and ncd motor domains on microtubules. *Proc. Natl Acad. Sci. USA* **93**, 9539–9544 (1996).
- Amos, L. A. & Hirose, K. The structure of microtubule-motor complexes. *Curr. Opin. Cell Biol.* **9**, 4–11 (1997).
- Berliner, E., Young, E. C., Anderson, K., Mahtani, H. K. & Gelles, J. Failure of a single-headed kinesin to track parallel to microtubule protofilaments. *Nature* **373**, 718–721 (1995).
- Vale, R. D. et al. Direct observation of single kinesin molecules moving along microtubules. *Nature* **380**, 451–453 (1996).

Acknowledgements. We thank S. Endow for the GST–MC5 clone; M. McNiven for the MMR48 antibody; M. Fritz for the rhodamine-labelled tubulin; U. Euteneuer, H. Felgner, R. Gräf and S. Linder for advice; and S. Fuchs for technical assistance. This work was supported by a Graduiertenkolleg fellowship to U.H. and a grant from the DFG to M.S.

Correspondence and requests for materials should be addressed to M.S. (e-mail: schliwa@bio.med.uni-muenchen.de).

Structure of the inhibitory receptor for human natural killer cells resembles haematopoietic receptors

Qing R. Fan^{*†}, Lidia Mosyak^{*†}, Christine C. Winter[‡], Nicolai Wagtmann[‡], Eric O. Long[‡] & Don C. Wiley^{*}

^{*} Department of Molecular and Cellular Biology, Howard Hughes Medical Institute, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

[‡] Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 12441 Parklawn Drive, Rockville, Maryland 20852, USA

[†] These authors contributed equally to this work.

Abnormal cells deficient in class I major histocompatibility complex (MHC) expression are lysed by a class of lymphocyte called natural killer (NK) cells¹. This lysis provides a defence against pathogens and tumour cells that downregulate MHC expression to avoid an MHC-restricted, T-cell immune response. Normal cells escape lysis because their MHC molecules are recognized by NK-cell inhibitory receptors, which inhibit lysis². Several such inhibitory receptor families have been described in humans and mice (reviewed in ref. 2). In the human killer-cell inhibitory receptor family, individual p58 members are specific for a subset of class I human leukocyte antigen (HLA)-C molecules. The human p58 natural killer-cell inhibitory receptor clone 42 recognizes HLA-Cw4, -Cw2 and -Cw6, but not HLA-Cw3, -Cw2, -Cw7 or -Cw8, which are recognized by p58 killer-cell inhibitor receptor clone 43 (ref. 3). We have determined the X-ray structure of the p58 NK-cell inhibitory receptor clone 42 at 1.7-Å resolution. The structure has tandem immunoglobulin-like domains positioned at an acute, 60-degree angle. Loops on the outside of the elbow between the domains form a binding site projected away from the NK-cell surface. The topology of the domains and their arrangement relative to each other reveal a

relationship to the haematopoietic receptor family, with implications for the signalling mechanism in NK cells.

The soluble ectodomain of the p58 natural killer-cell inhibitory receptor (KIR) that we crystallized was expressed in bacteria and refolded *in vitro*⁴. It specifically blocks the binding of a KIR–immunoglobulin fusion protein to cells expressing HLA-Cw4, but has no effect on the binding of a KIR–immunoglobulin fusion protein specific for HLA-CW3 to cells expressing HLA-CW3 (ref. 4). The entire KIR ectodomain, residues 1–224, and a shorter form, residues 1–200, without the stem domain of 24 residues, were both shown by a gel-shift assay to bind to soluble, recombinant HLA-Cw4 (ref. 4). As predicted from sequence similarity⁵, the overall fold of each KIR domain resembles immunoglobulin-like domains containing two antiparallel β -sheets (Fig. 1a, b). Topologically, the p58 KIR structure has tandem fibronectin type III-like domains, which are structurally similar to immunoglobulin domains. The two domains are equal in size (domain D1, residues 6–101; domain D2, residues 105–200), and are connected by a three-residue linker (residues 102–104) (Fig. 1b).

The topology of the D1 immunoglobulin-like domain is of the h-type⁶, similar to that of the D1 domains of the human growth hormone receptor (hGHR)⁷, human prolactin receptor (hPLR)⁷, and the erythropoietin receptor (EPOR)⁸. The D2 domain has the closely related s-type topology⁶ found in the D2 domains of hGHR, hPLR and EPOR^{8–10}, and both domains of class II receptors of the haematopoietic superfamily, including the α -chain of the interferon- γ receptor (IFN- γ R α)¹¹, and tissue factor^{12,13}. The s-type topology has also been found in other two-domain structures, including the co-receptor CD4 (refs 14, 15), the cell-adhesion molecule CD2 (ref. 16) and neuroglian¹⁷. However, in both CD2 and CD4, the s-type domain is paired in tandem with a variable-type immunoglobulin domain and the two domains assume a linear conformation, lacking the elbow region found in all of the haematopoietic receptors and KIR. Although some members of the haematopoietic receptor superfamily contain additional domains, ligand binding is mediated through the h- or s-type domains⁷. The s-type topology of D2 has a β -sheet of three strands, A, B and E (Fig. 1a, b, dark orange) packed against a β -sheet of four strands, C', C, F and G (Fig. 1a, b, light orange). In h-type topology of the D1 domain, the C' strand is elongated, pairing first with the C strand of the C'CFG β -sheet and at its far end (where it is renamed D) pairing with the E strand of the ABE(D) β -sheet (Fig. 1a, b). The switch from one β -sheet to the other occurs at a kink in strand C'/D at Gly 53. When superimposed, the D1 and D2 domains, which have 40% sequence identity, are strikingly similar in structure with a root-mean-square (r.m.s.) deviation of 0.9 Å for 85 C α pairs.

Table 1 Statistics for data collection and refinement

Data collection								
Data set	Wavelength (Å)	Scattering factors (e)		Resolution (Å)	Unique reflections	Average $I/\sigma(I)$	Completeness (%)	R_{sym} (%)†
		f'	f''					
MAD	0.9639 (remote)	–3.49	3.72	10.0–2.2	11,480	18.9	98.7 (99.1)	7.3 (11.7)
	0.9791 (peak)	–7.89	4.72	10.0–2.2	11,449	18.5	98.6 (99.1)	7.2 (11.6)
	0.9794 (edge)	–9.86	3.05	10.0–2.2	11,495	18.8	98.8 (99.1)	7.0 (11.3)
native	0.918			16.0–1.7	25,065	171	99.5 (97.8)	7.3 (23.4)
Refinement (6–1.7 Å)								
R_{cryst} (R_{free})‡ (%)	Reflections (free)	Non-hydrogen protein atoms		solvent molecules	R.m.s. deviations			B -factors (Å ²)
					Bonds (Å)	Angles (°)		
20.5 (25.5)	24,473 (2,427)	1,513		211	0.010	1.83		2.85

Values in parentheses correspond to the last resolution shell; 2.28–2.20 Å for MAD data sets, or 1.76–1.70 Å for native data set.

* The values of f' and f'' for λ_{peak} and λ_{edge} were derived from the experimental values of the absorption spectrum; values for λ_{remote} were calculated from theoretical cross-sections³⁰.

† $R_{\text{sym}} = \sum |I_i - \langle I \rangle| / \sum I_i$, where I_i is the intensity of an individual reflection, and $\langle I \rangle$ is the average intensity of that reflection. Bijvoet mates were considered as separate reflections.

‡ $R_{\text{cryst}} = \sum ||F_o| - |F_c|| / \sum |F_o|$, where F_c is the calculated structure factor. R_{free} is equivalent to R_{cryst} , but calculated for a randomly chosen 10% of reflections that were omitted from the refinement process.

KIR domains define a new topological subtype⁶, which we call Kh- and Ks-type, in which the carboxy-terminal half of strand A, named A', pairs with strand G to form a five-stranded sheet C'CFA' (Fig. 1a, b, light orange) by switching at *cis* prolines (residues 14 and 114) from one β -sheet to another, as found in variable domains of antibody light chains, CD8 and T-cell antigen receptors.

A signature sequence motif of class I haematopoietic receptors, the WSXWS box¹⁸, is also present in p58 KIR. The p58 clone 42 KIR (p58-cl42) contains at residues 188–192 a variant of the WSXWS sequence motif (WSKSS) in D2, in the same location as found in the haematopoietic receptor family (Fig. 1a, b). Like the motif in the other structures, the WSKSS sequence in KIR assumes a polyproline type II conformation, resulting in a wide β -bulge^{8–10} in which the hydroxyl groups of the two conserved serine residues (underlined) form hydrogen bonds to the adjacent F-strand backbone, replacing the usual backbone-to-backbone hydrogen bonds between strands in a β -sheet. A surface cluster of aromatic and charged side chains near the motif present in several haematopoietic receptors^{8–10} is absent in KIR. KIR differs from members of the haematopoietic receptor superfamily in that a variant of the motif also occurs in D1 (Fig. 1a, b). The sequence VSAPS, residues 90–94, forms the same type of β -bulge at the homologous position. Variants of both the VSAPS and WSKSS sequences are found in the sequences of all KIR

immunoglobulin-like domains, with the two structurally important serines completely conserved.

Helical linkers of at least four residues in length have been found linking the D1 and D2 domains in five members of the haematopoietic receptor family: hGHR, hPLR, EPOR, IFN- γ R α and tissue factor^{8–13}. The D1 and D2 domains of the KIR are more tightly connected by a three-residue linker (residues 102–104) with a helical conformation ($\phi = -95.7$ deg, $\psi = -18.4$ deg) at Leu 104.

Most of the side chains in the D1–D2 interface are conserved in all KIR sequences, and the few differences there are appear consistent with the observed packing, suggesting that the V shape will be a feature of all p58 and p70 KIRs. Extensive hydrogen bonding between strands of both domains and the large buried solvent-accessible surface area, 1,076 Å² (calculated using the program SURFACE¹⁹; probe radius 1.4 Å), suggest that the elbow angle is fixed. The D1–D2 interface includes a three-stranded parallel β -sheet formed by the two strands A' and G from the D1 domain and a short segment of strand G in the D2 domain (Fig. 1a, b). D1–D2 contacts occur not only at the tip of the V (Fig. 2a, black) but also across the gap between the arms of the V (Fig. 2a, red). Most of the polar interactions found at the interface involve main-chain atoms, and are therefore independent of sequence variation.

Site-directed mutations indicate strongly that the binding site on

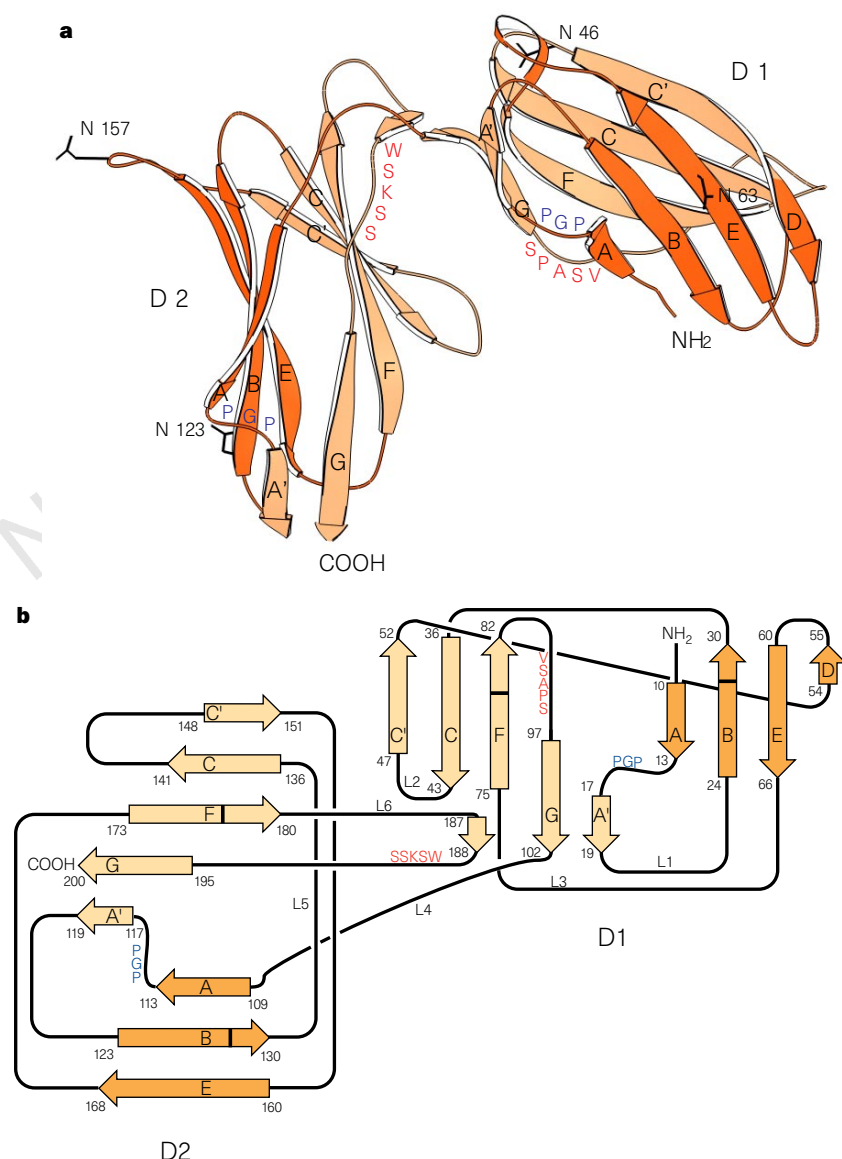


Figure 1 Structure and topology of p58-cl42 KIR. **a**, KIR domain D1 is N-terminal, and D2 is C-terminal; ABED β -sheet (dark orange) and C'CFA' β -sheet (light orange); WSKSS and VSAPS motifs (red); potential glycosylation sites at asparagines 46, 63, 123 and 157 (black). **b**, Topological diagram showing the h-type immunoglobulin-like fold of D1 and the s-type fold of D2. The β -sheet at the D1–D2 interface (strands A' and G in D1 and a fragment of strand G in D2) is shown below D1. Loops L1–L6 are implicated in ligand binding by the haematopoietic receptors, hGHR, hPLR and EPOR. Cysteines 28, 79, 128 and 177 are indicated by black bars.

indicates that the signalling mechanism, at least of that KIR, may involve molecules other than a putative receptor dimer. The stem region of KIR may contribute to HLA binding (Q.R.F. *et al.*, unpublished data). The full ectodomain of p58-cl42 KIR (residues 1–224) completely blocked the binding of the shorter form (residues 1–200 lacking the stem segment) to soluble HLA-Cw4 on band-shift assays (data not shown). Furthermore, binding of the shorter form to HLA-Cw4 was reversed upon addition of the longer form, suggesting a higher dissociation constant for the complex between HLA-Cw4 and the shorter form. The more favourable binding observed between HLA-Cw4 and the KIR that includes the stem (residues 220–224) suggests a role for the C-terminal stem in ligand binding, possibly by stabilizing dimerization (Fig. 3e), considering the distance between the C terminus and the ligand-binding site (Fig. 2b).

The second receptor polypeptide chain that we postulate may bind in either a homodimer or heterodimer to a KIR–MHC complex (Fig. 3e) is expected, by analogy to the haematopoietic receptors: to have a much lower affinity for HLA, and bind only after

the first KIR has bound; to bind to a second, as-yet undiscovered, site on MHC molecules; and to interact with the first KIR polypeptide at some as-yet undiscovered site, probably in the D2 domain. In the haematopoietic receptors, the affinity of the second receptor chain to the single-receptor/ligand complex is 500- to 1,000-fold weaker than the first, making detection and characterization of the second binding sites difficult⁷. This may account for the current lack of evidence for KIR dimerization by MHC molecules.

Methods

Incorporation of Se-Met. The selenomethionyl (Se-Met) p58-cl42 KIR (residues 1–200) was expressed in *Escherichia coli* as inclusion bodies, refolded by dilution, and purified following the same procedure as native p58-cl42 (ref. 4). The incorporation of Se-Met was achieved by growing the cells in minimal medium (M9) supplemented with 2 mM MgSO₄, 0.1 mM CaCl₂, 0.2% glucose, 0.00005% thiamine, and 40 mg l⁻¹ each of all L-amino acids except Cys and Met, and inducing the cells in the presence of 120 mg l⁻¹ of Se-Met along with 100 mg l⁻¹ each of Thr, Lys and Phe, and 50 mg l⁻¹ each of Leu, Ile and Val. All

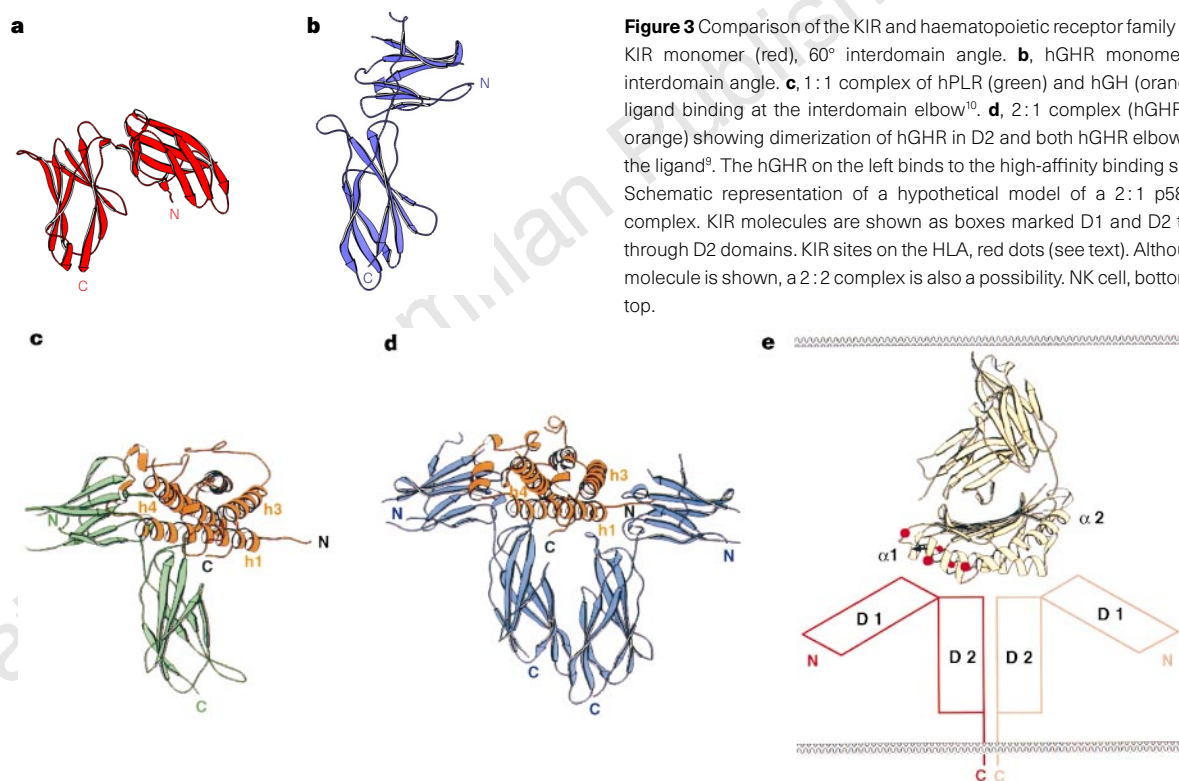


Figure 3 Comparison of the KIR and haematopoietic receptor family structures. **a**, KIR monomer (red), 60° interdomain angle. **b**, hGHR monomer (blue), 90° interdomain angle. **c**, 1:1 complex of hPLR (green) and hGH (orange), showing ligand binding at the interdomain elbow¹⁰. **d**, 2:1 complex (hGHR, blue; hGH, orange) showing dimerization of hGHR in D2 and both hGHR elbows contacting the ligand⁹. **e**, Schematic representation of a hypothetical model of a 2:1 p58 KIR/HLA-C complex. KIR molecules are shown as boxes marked D1 and D2 that dimerize through D2 domains. KIR sites on the HLA, red dots (see text). Although one HLA molecule is shown, a 2:2 complex is also a possibility. NK cell, bottom; target cell, top.

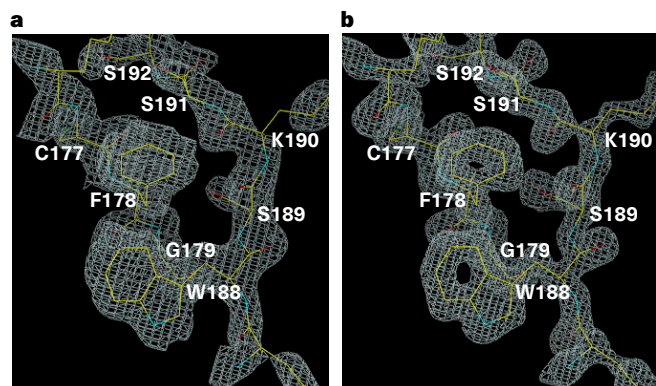


Figure 4 Electron-density maps showing the refined model in the region of the WSKSS sequence. **a**, A 2.2-Å resolution Fourier map calculated with MAD phases improved by density modification. **b**, A 1.7-Å resolution simulated-annealing-omit map calculated with the $2F_{\text{obs}} - F_{\text{calc}}$ amplitudes and model phases. Both maps are contoured at 1.2σ level.

five Met were substituted with Se-Met as determined by mass spectrometry.

Crystallization. Crystals of p58-cl42 were grown from 0.55 M $(\text{NH}_4)_2\text{HPO}_4$, 50 mM sodium citrate, pH 5.4, final pH 7.7. Hexagonal rods ($0.1 \times 0.1 \times 1.5 \text{ mm}^3$; space group $P6_1$; $a = b = 92.4 \text{ \AA}$, $c = 46.8 \text{ \AA}$; 1 molecule per asymmetric unit) grow two weeks after seeding. Se-met p58 crystallized under the same condition using the native crystals as seeds; the difference in cell dimensions was less than 0.5%. After the crystals were stabilized for at least 10 h in a harvesting solution (1.5 M $(\text{NH}_4)_2\text{HPO}_4$, 50 mM sodium citrate, pH 5.4, final pH 7.7), they were soaked for 2–5 min in a cryo-protecting solution (1.5 M $(\text{NH}_4)_2\text{HPO}_4$, 50 mM Na citrate, pH 5.4, 25% glycerol, final pH 7.7), and flash-cooled with liquid nitrogen.

Data collection. Multiwavelength anomalous dispersion (MAD)²⁶ data were collected to 2.2 Å with a 300-mm diameter MAR Research image-plate system at the X25 beamline of the National Synchrotron Light Source, Brookhaven National Laboratory. A high-resolution native data set was collected to 1.7 Å on the Princeton 2K CCD detector at F-1 beamline of the Cornell High Energy Synchrotron Source (CHESS). Data were processed (Table 1) using DENZO and SCALEPACK (HKL Research). Most of the subsequent processing used the CCP4 programs¹⁹.

MAD phasing. MAD phasing was treated as a case of multiple isomorphous replacement²⁷. Four selenium sites were identified from anomalous and dispersive difference Pattersons and were checked by difference Fourier. The N-terminal methionine was disordered. Refinement of anomalous scatterer parameters and phase calculation were performed with MLPHARE²⁸. Because of discrepancies between phasing statistics generated by MLPHARE and other programs²⁷, electron-density maps before and after model refinement are displayed instead of experimental phasing statistics (Fig. 4). The initial MAD map was improved by density modification using DM¹⁹, assuming 40% solvent content. The correct space-group enantiomer $P6_1$ was identified by the presence of clear solvent boundary in the 2.2-Å electron-density maps.

Model refinement. The experimental MAD phases were used with the native data set to calculate the electron density for the native structure. Both density-modified and unmodified electron-density maps were used to build an 85% complete model with O (DATAON AB). For refinement, data with $|F_{\text{obs}}| > 0$ were included. The model was initially refined at 10–2.2 Å using positional refinement and simulated annealing protocols in X-PLOR²⁹. Several cycles of manual refitting and subsequent inclusion of lower-resolution data to 1.6 Å combined with bulk solvent correction allowed the missing loop regions to be traced. The resolution was then extended in one step to 1.7 Å. Refinement at this stage involved simulated annealing followed by B-factor refinement, with the extensive use of simulated annealing omit maps (Fig. 4b). The final model refined in X-PLOR contained residues 6–200 and 211 water molecules. This model was refined with REFMAC¹⁹ (Table 1). The maximum-likelihood method in REFMAC lowered the R-values in the highest resolution shell (from 38.4 to 33.7% for R_{free} at 1.76–1.7 Å). All ϕ and ψ angles lie in the allowed regions of the Ramachandran plot, with 92% in the most favourable regions. Side-chain densities are well defined for all residues except 151–153 in a loop, which have B-factors $> 70 \text{ \AA}^2$.

Figure preparation. Figures 1a, 2 and 3 were prepared using the program RIBBONS³⁰.

Received 27 May; accepted 3 July 1997.

1. Ljunggren, H. G. & Karre, D. In search of the "missing self": MHC molecules and NK cell recognition. *Immunol. Today* **11**, 237–244 (1990).
2. Lanier, L. L. Natural killer cells: from no receptors to too many. *Immunity* **6**, 371–378 (1997).
3. Long, E. O. et al. Killer cell inhibitory receptors: diversity, specificity, and function. *Immunol. Rev.* **155**, 135–144 (1997).
4. Fan, Q. R. et al. Direct binding of a soluble natural killer cell inhibitory receptor to a soluble human leukocyte antigen-Cw4 class I major histocompatibility complex molecule. *Proc. Natl Acad. Sci. USA* **93**, 7178–7183 (1996).
5. Wagtman, N. et al. Molecular clones of the p58 natural killer cell receptor reveal Ig-related molecules with diversity in both the extra- and intracellular domains. *Immunity* **2**, 439–449 (1995).
6. Bork, P., Holm, L. & Sander, C. The immunoglobulin fold: structural classification, sequence patterns and common core. *J. Mol. Biol.* **242**, 309–320 (1994).
7. Kossiakoff, A. A. et al. Comparison of the intermediate complexes of human growth hormone bound to the human growth hormone and prolactin receptors. *Protein Sci.* **3**, 1697–1705 (1994).
8. Livnah, O. et al. Functional mimicry of a protein hormone by a peptide agonist: the EPO receptor complex at 2.8 Å. *Science* **273**, 464–471 (1996).
9. De Vos, A. M., Ultsch, M. & Kossiakoff, A. A. Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* **255**, 306–312 (1992).
10. Somers, W., Ultsch, M., De Vos, A. M. & Kossiakoff, A. A. The X-ray structure of a growth hormone–prolactin receptor complex. *Nature* **372**, 478–481 (1994).
11. Walter, M. R. et al. Crystal structure of a complex between interferon- γ and its soluble high-affinity

receptor. *Nature* **376**, 230–235 (1995).

12. Müller, Y. A., Ultsch, M. H., Kelly, R. F. & De Vos, A. M. Structure of the extracellular domain of human tissue factor: location of the factor VIIa binding site. *Biochemistry* **33**, 10864–10870 (1994).
13. Harlos, K. et al. Crystal structure of the extracellular region of human tissue factor. *Nature* **370**, 662–666 (1994).
14. Wang, J. et al. Atomic structure of a fragment of human D4 containing two immunoglobulin-like domains. *Nature* **348**, 411–418 (1990).
15. Ryu, S. et al. Crystal structure of an HIV-binding recombinant fragment of human CD4. *Nature* **348**, 419–426 (1990).
16. Jones, E. Y., Davis, S. J., Williams, A. F., Harlos, K. & Stuart, D. I. Crystal structure at 2.8 Å resolution of a soluble form of the cell adhesion molecule CD2. *Nature* **360**, 232–239 (1992).
17. Huber, A. H., Wang, Y. E., Bieber, A. J. & Bjorkman, P. J. Crystal structure of tandem type III fibronectin domains from *Drosophila* neuroglian at 2.0 Å. *Neuron* **12**, 717–731 (1994).
18. Bazan, J. F. Structural design and molecular evolution of a cytokine receptor family. *Proc. Natl Acad. Sci. USA* **87**, 6934–6938 (1990).
19. Collaborative computational project No. 4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–776 (1994).
20. Winter, C. C. & Long, E. O. A single amino acid in the p58 killer cell inhibitory receptor controls the ability of NK cells to discriminate between the two groups of HLA-C-Allotypes. *J. Immunol.* **158**, 4026–4028 (1997).
21. Burshtyn, D. N. et al. Recruitment of tyrosine phosphatase HCP by the killer cell inhibitory receptor. *Immunity* **4**, 77–85 (1996).
22. Binstadt, B. A. et al. Sequential involvement of Lck and SHP-1 with MHC-recognizing receptors on NK cells inhibits Fc γ -initiated tyrosine kinase activation. *Immunity* **5**, 629–638 (1996).
23. Ihle, J. N., Witthuhn, B. A., Quelle, F. W., Yamamoto, K. & Silvennoinen, O. Signaling through the hematopoietic cytokine receptors. *Annu. Rev. Immunol.* **13**, 369–398 (1995).
24. Biassoni, R. et al. The human leukocyte antigen (HLA)-C-specific "activatory" or "inhibitory" natural killer cell receptors display highly homologous extracellular domains but differ in their transmembrane and intracytoplasmic portions. *J. Exp. Med.* **183**, 645–650 (1996).
25. Pende, D. et al. The natural killer cell receptor specific for HLA-A allotypes: a novel member of the p58/p70 family of inhibitory receptors that is characterized by three immunoglobulin-like domains and is expressed as a 140-kD disulfide-linked dimer. *J. Exp. Med.* **184**, 505–518 (1996).
26. Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58 (1991).
27. Ramakrishnan, V. & Biou, V. Treatment of MAD data as a special case of MIR. *Meth. Enzymol.* **276**, 538–557 (1996).
28. Otwinowski, Z. *Isomorphous Replacement and Anomalous Scattering* (Daresbury Laboratory, UK, 1991).
29. Brünger, A. T. *X-PLOR, Version 3.1: A System for X-ray and NMR* (Yale Univ. Press, New Haven, CT, 1992).
30. Carson, M. Ribbon models of macromolecules. *J. Mol. Graph.* **5**, 103–106 (1987).
31. Cromer, D. T. Calculation of anomalous scattering factors at arbitrary wavelengths. *J. Appl. Crystallogr.* **16**, 437 (1983).

Acknowledgements. We thank D. N. Garboczi, X. D. Zhang, P. Rosenthal, K. Smith and L. Chen for help with data collection; R. Hellmiss for preparing Fig. 1b; W. I. Weiss and V. Ramakrishnan for discussion; Y. Liu for the program for calculating f' and f'' values; W. Kossiakoff for the coordinates of the hPLR–hGH complex; Harvard microchemistry facility for mass spectroscopy; A. Haykov and R. Crouse for technical support; the staff at the Brookhaven X25 beamline and CHESS for assistance with data collection; and D. N. Garboczi for reading the manuscript. Q.R.F. is a recipient of an NSF predoctoral fellowship. L. M. is an associate and D. C. W. is an investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to D.C.W. (e-mail: wiley@crystal.harvard.edu). Coordinates will be deposited in the Brookhaven Data base and are available now from Q.R.F. (e-mail: fan@crystal.harvard.edu).

erratum

Neurotactin, a membrane-anchored chemokine upregulated in brain inflammation

Yang Pan, Clare Lloyd, Hong Zhou, Sylvia Dolich, Jim Deeds, Jose-Angel Gonzalo, Jim Vath, Mike Gosselin, Jingya Ma, Barry Dussault, Elizabeth Woolf, Geoff Alperin, Janice Culpepper, Jose Carlos Gutierrez-Ramos & David Gearing

Nature **387**, 611–617 (1997)

The sequence of neurotactin reported in this Letter is identical to that of fractalkine, a CX₃C membrane-bound chemokine reported in a Letter by J. F. Bazan *et al.* in *Nature* **385**, 640–644 (1997), published while the paper by Pan *et al.* was under consideration. A note to this effect in the paper by Pan *et al.* was inadvertently omitted by *Nature*. □

Keeping growth under control

Improved technology for incubating, processing and observing tissues and cells continues to develop, and includes a transportable submarine incubation system and semi-enclosed, enclosed and insulated tissue processors.

PHD Harvesters

From Semat

For work in all phases of sample preparation, this system helps speed throughput for a wide range of tests from MLC to RBA

Made from stainless steel and aluminium, the rugged and durable harvesters will cut samples out of filter material and deposit them directly into any type of small scintillation vial or test tube. This is said to eliminate the need to handle individual samples, prevents mix-ups and provides better operator protection from high-risk materials. With 24 samples harvested and deposited each time, up to 1.3 million lymphocytes or 2.5 mg of homogenized tissue can be harvested per sample, allowing a 96-well plate to be processed and prepared for scintillation cocktail in just a few minutes. Many accessories and options are available and can be retrofitted, including special harvest heads for use with test tubes, 24- and 48-well plates, multiple wash timers, wash pumps and supernatant collection systems.

Reader Enquiry No. 100

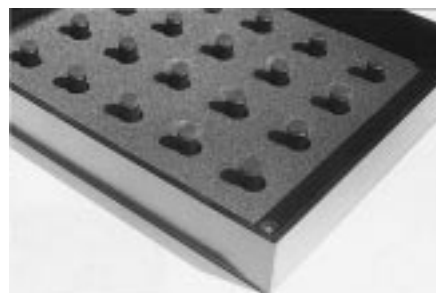
Slide Show

From Genex

A range of incubation boxes for improved handling of microscope slides and cover slips for humid incubations in histology or tissue culture laboratories

The new range comprises purpose-built humid incubation boxes that grip the slides in place, as well as elevated rails or pads coated with a material that sticks to the underside of glass to prevent unwanted movement, but also allow easy removal of the slide. One model has a capacity for 32 slides, whereas the other provides for 24 coverslips. Each has a neat stackable design and a choice of either clear or black (fluorescence) lid. The boxes are made from black PVC and acrylic for robust structure and easy cleaning. A washable foam insert is supplied to help maintain the humid environment.

Reader Enquiry No. 101



The Genex 24 box for tissue culture and histology.



Merck offers a range of slides for formalin-fixed frozen tissue sections.

SuperFrost Plus Gold

From Merck

Microscope slides that have been designed for fresh or formalin-fixed frozen tissue sections

These slides are made with a new glass-adhesive technology, which first attracts and then chemically bonds frozen tissue sections firmly to the surface of the slide. This is particularly useful for tissue sections that are normally difficult to bond, such as breast, brain or skin samples. The slides are useful with special stains, and in immunocytochemical and *in situ* hybridization techniques. Moreover, the use of these slides is said to eliminate the need to add spray or adhesives to the slide. Merck also offers Menzel diagnostic slides with 10 standard executions and Teflon or epoxy finishes available. The slides' printed coating withstands test procedures, solvents and multiple sterilizations — ground edges allow for safer handling. Merck's Superslips help automatic coverslippers work the way in which they were designed. They have moisture resistance built into the glass and desiccants in the package. This combination helps eliminate most moisture-related machine shut downs.

Reader Enquiry No. 102

TSIS

From Embryo Technology Center

The transportable submarine incubation system (TSIS) is for culturing tissues, cells or embryos

Culture dishes are individually wrapped into a laminated foil that is not permeable to CO₂, O₂ and N₂, which is then filled with the appropriate gas mixture, heat-sealed and submerged into a transportable, temperature-controlled water bath that has been adjusted to the proper temperature. Samples can then be incubated for the required period (up to eight days). Thus, each culture dish exists as an individual incubator and

the mediator of heat is water instead of air. This system is said to provide improved stability of temperature, humidity and gas mixture; quick recovery of parameters after opening; the flexibility to use different gas mixtures; safety in operation; as well as cost efficiency. It is also well suited for establishing and maintaining appropriate culture conditions under field conditions.

Reader Enquiry No. 103

ImmobaSil

From Integra Biosciences

Support for high-density cell culture

This microporous silicone rubber support is stated to provide up to 50 per cent higher densities of attachment-dependent cells. Both non-toxic and biocompatible with microbial, plant and mammalian cells, this support has an effective colonization area of 50,000 m² m⁻³, according to the manufacturer. For cells in suspension, 90 per cent entrapment, even in stirred systems, can be reached. Furthermore, improved oxygen mass transfer to immobilized cells allows reduced prefixing times, minimizes cell damage caused by shear and reduces power costs. ImmobaSil is available in four formats: FS (1-mm diam. × 0.25-mm length) for tissue culture increased surface-to-volume ratios; ImmobaSil G (1-mm diam. × 1-mm length), a general-purpose microcarrier that can be weighted for fluidized bed applications; ImmobaSil R (a 4-mm o.d. × 2-mm i.d. × 4-mm Raschig ring) for minimal flow resistance in packed-bed reactors; and ImmobaSil D (a 10-mm diam. × 1-mm disc) with a planar geometry that maximizes surface-to-volume ratios and maintains hydrodynamic stability. For use in high-stress systems, this support is resistant to

ADVERTISEMENT

MMBG 5x1

both biological and chemical attack to help ensure long life and easy cleaning. In addition, it can withstand a wet heat of up to 200 °C, making it compatible with standard steam sterilization and autoclave processes.

Reader Enquiry No. 104

Fibra-Cel

From Bibby Sterilin

A microcarrier material for a wide range of anchorage-dependent cell culture applications

The carriers consist of 6-mm discs of non-woven polyester, each bound to a polypropylene screen, which acts as a weighting device. The surfaces are tissue culture treated to provide conditions for maximum cell attachment and growth. They form a dramatically enlarged surface area for high-volume production of biological macromolecules. Fibra-Cel is stated to facilitate cell growth to densities in excess of $5 \times 10^7 \text{ ml}^{-1}$. This can be achieved with very low cell seeding densities, down to as little as a stated one per cent of the final density. The material should protect fragile cells from shear damage, by allowing growth within the fibre disc. For packed-cell continuous cultures, the discs can be immobilized in a compartment and perfused with nutrient medium. Product secreted into the medium can then be harvested, cell-free, for downstream processing and purification. In a stirred system, the discs can be mixed and will settle to allow harvesting or removal of the supernatant; in a roller system, the surface area for cell growth can be increased by incorporating the discs within the roller bottle.

Reader Enquiry No. 105

Clinicut 60 workstation

From Bright

A specimen freezer that not only has a 'cold spot' inside the chamber, but is an insulated workstation for tissue preparation

The system has a convenient specimen holder and uses inexpensive specimen discs that fit into a quick-release holder. The microtome component is designed for trimming and reliable sectioning, and has a radial, flat cutting action. All microtome controls are readily accessible and integral sliding guards contribute to operator safety. Furthermore, the microtome can be removed quickly without using tools. The unit has an easy-to-clean chamber and uses a one-step method to remove or install the microtome. 'Automatic defrost' ensures efficient refrigeration and convenience of use. The Quick Freezer operates continuously between -44 °C and -53 °C (at a chamber temperature of -20 °C), enabling the user to prepare specimens in various ways, for example, using the embedding moulds, pre-shaped blocks can be made ready for cutting. Alternatively, the user can freeze on the flat area in the conven-

tional way. The unit is fully insulated and equipped with a cover to prevent frosting, and the freezer plates are easily removed for cleaning. Tissue trimmings are kept out of the microtome mechanism by the cover and a debris screen, and are collected in the removable debris trays. The freezer plate is easy to take out for cleaning and the system has no hidden corners to clean.

Reader Enquiry No. 106

TP 1020 tissue processor

From Leica

A semi-enclosed automatic tissue processor for the histology laboratory

The processor features a carousel design with twelve 1.8-l stations, ten reagent stations, two wax baths and a standard metal tissue basket with an 80-cassette capacity. Nine programs can be run with immediate or delayed start — the delayed start functions for up to nine days after programming. The system also allows for the interruption of an automatic process in special applications where reloading or removing cassettes is necessary before the end of a run. Tissue specimens are protected from drying out even during a power failure as the tissue baskets are automatically immersed in a station. Safe lids with rubber seals reduce exposure to hazardous fumes. Moreover, there are several configurations available, such as two tissue basket operation, vacuum function, and a fume control system.

Reader Enquiry No. 107

Pathcentre Extensions

From Shandon

Extensions for the company's enclosed tissue processing units have all the features of a standard unit, but without the command console

The extension units allow extra processing capacity to be added at any time without buying a completely new system. Up to three units can be added to a single Pathcentre processor, increasing the total cassette capacity from 324 to 1,296 cassettes. As many as nine processing routines can be programmed and memorized per unit — with add-on modules the capacity can be increased to a total of 36 programs. Thus, individual programs can be optimized for different specimen types. A new screen display allows the user to monitor the status of all four modules at a glance. This is a totally enclosed tissue processor that does not vent any of the fumes associated with conventional tissue processing into the laboratory during processing. The circular shape of the reaction chamber in combination with special agitation systems helps to optimize tissue processing efficiency and to prolong reagent life, says the manufacturer.

Reader Enquiry No. 108

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

bio1016x1

HTI 7x1

dianorm 7x1